A probabilistic study of neural complexity

Jerome Buzzi, Lorenzo Zambotti

A PROBABILISTIC STUDY OF NEURAL COMPLEXITY

J. BUZZI AND L. ZAMBOTTI

Abstract. G. Edelman, O. Sporns, and G. Tononi have introduced the *neural complexity* of a family of random variables, defining it as a specific average of mutual information over subfamilies. We show that their choice of weights satisfies two natural properties, namely exchangeability and additivity, and we call any functional satisfying these two properties an *intricacy*. We classify all intricacies in terms of probability laws on the unit interval and study the growth rate of maximal intricacies when the size of the system goes to infinity. For systems of a fixed size, we show that maximizers have small support and exchangeable systems have small intricacy. In particular, maximizing intricacy leads to spontaneous symmetry breaking and failure of uniqueness.

1. Introduction

1.1. **A functional over random systems.** Natural sciences have to deal with "complex systems" in some obvious and not so obvious meanings. Such notions first appeared in thermodynamics. Entropy is now recognized as the fundamental measure of complexity in the sense of randomness and it is playing a key role as well in information theory, probability and dynamics [12]. Much more recently, subtler forms of complexity have been considered in various physical problems [1, 3, 7, 11], though there does not seem to be a single satisfactory measure yet.

Related questions also arise in biology. In their study of high-level neural networks, G. Edelman, O. Sporns and G. Tononi have argued that the relevant complexity should be a combination of high *integration* and high *differentiation*. In [22] they have introduced a quantitative measure of this kind of complexity under the name of *neural complexity*. As we shall see, this concept is strikingly general and has interesting mathematical properties.

In the biological [10, 13, 14, 16, 17, 18, 19, 20, 23, 24] and physical [2, 8] literature, several authors have used numerical experiments based on Gaussian approximations and simple examples to suggest that high values of this neural complexity are indeed associated with non-trivial organization of the network, away both from complete disorder (maximal entropy and independence of the neurons) and complete order (zero entropy, i.e., complete determinacy).

2000 *Mathematics Subject Classification.* 94A17, 92B30, 60C05.

Key words and phrases. Entropy; Mutual information; Complexity; Discrete Probability; Exchangeable vectors.

The aim of this paper is to provide a mathematical foundation for the Edelman-Sporns-Tononi complexity. Indeed, it turns out to belong to a natural class of functionals: the *averages of mutual informations satisfying exchangeability and weak-additivity* (see below and the Appendix for the needed facts of information theory). The former property means that the functional is invariant under permutations of the system. The latter that it is additive over independent systems. We call these functionals *intricacies* and give a unified probabilistic representation of them.

One of the main thrusts of the above-mentioned work is to understand how systems with large neural complexity look like. From a mathematical point of view, this translates into the study of the maximization of such functionals (under appropriate constraints).

This maximization problem is interesting because of the trade-off between high entropy and strong dependence which are both required for large mutual information. Such *frustration* occurs in spin glass theory [21] and leads to asymmetric and non-unique maximizers. However, contrarily to that problem, our functional is completely deterministic and the symmetry breaking (in the language of theoretical physics) occurs in the maximization itself: we show that the maximizers are not exchangeable although the functional is. We also estimate the growth of the maximal intricacy of finite systems with size going to infinity and the size of the support of maximizers.

The computation of the exact growth rate of the intricacy as a function of the size and the analysis of systems with almost maximal intricacies build on the techniques of this paper, especially the probabilistic representation below, but require additional ideas, so are deferred to another paper [5].

1.2. **Intricacy.** We recall that the *entropy* of a random variable X taking values in a finite or countable space E is defined by

$$\mathrm{H}(X) := -\sum_{x\in E} P_X(x)\,\log(P_X(x)), \qquad P_X(x) := \mathbb{P}(X = x).$$

Given two discrete random variables defined over the same probability space, the *mutual information* between X and Y is

$$\mathrm{MI}(X, Y) := \mathrm{H}(X) + \mathrm{H}(Y) - \mathrm{H}(X, Y).$$

We refer to the appendix for a review of the main properties of the entropy and the mutual information and to [6] and [12] for introductions to information theory and to the various roles of entropy in mathematical physics, respectively. For now, it suffices to recall that $\mathrm{MI}(X, Y) \geq 0$ is equal to zero if and only if X and Y are independent, and therefore $\mathrm{MI}(X, Y)$ is a measure of the dependence between X and Y.

Edelman, Sporns and Tononi [22] consider systems formed by a finite family $X = (X_i)_{i\in I}$ of random variables and define the following concept of complexity. For any $S \subset I$, they divide the system in two families

$$X_S := (X_i, i \in S), \qquad X_{S^c} := (X_i, i \in S^c),$$

where $S^c := I\backslash S$. Then they compute the mutual informations $\mathrm{MI}(X_S, X_{S^c})$ and consider an average of these:

$$\mathcal{I}(X) := \frac{1}{|I|+1} \sum_{S \subset I} \frac{1}{\binom{|I|}{|S|}} \, \mathrm{MI}(X_S, X_{S^c}), \tag{1.1}$$

where $|I|$ denotes the cardinality of I and $\binom{n}{k}$ is the binomial coefficient. Note that $\mathcal{I}(X)$ is really a function of the *law* of X and not of its random values.

The above formula can be read as the expectation of the mutual information between a random subsystem X_S and its complement X_{S^c} where one chooses uniformly the size $k \in \{0, \dots, |I|\}$ and then a subset $S \subset I$ of size $|S| = k$.

In this paper we prove that $\mathcal{I}$ fits into a natural class of functionals, which we call **intricacies**. We shall see that these functionals have very similar, though not identical properties and admit a natural and technically very useful probabilistic representation by means of a probability measure on $[0, 1]$.

Notice that $\mathcal{I} \geq 0$ and $\mathcal{I} = 0$ if and only if the system is an independent family (see Lemma 3.9 below). In particular, both complete order (a deterministic family X) and total disorder (an independent family) imply that every mutual information vanishes and therefore $\mathcal{I}(X) = 0$.

On the other hand, to make (1.1) large, X must simultaneously display two different behaviors: a non-trivial correlation between its subsytems and a large number of internal degrees of freedom. This is the hallmark of complexity according to Edelman, Sporns and Tononi. The need to strike a balance between *local independence* and *global dependence* makes such systems not so easy to build (see however Example 2.10 and Remark 2.11 below for a simple case). This is the main point of our work.

1.3. **Intricacies.** Throughout this paper, a **system** is a finite collection $(X_i)_{i \in I}$ of random variables, each X_i, $i \in I$, taking value in the same finite set, say $\{0, \dots, d-1\}$ with $d \geq 2$ given. Without loss of generality, we suppose that I is a subset of the positive integers or simply $\{1, \dots, N\}$. In this case it is convenient to write N for I.

We let $\mathcal{X}(d, I)$ be the set of such systems and $\mathcal{M}(d, I)$ the set of the corresponding laws, that is, all probability measures on $\{0, \dots, d-1\}^I$ for any finite subset I. We often identify it with $\mathcal{M}(d, N) := \mathcal{M}(d, \{1, \dots, N\})$ for $N = |I|$. If X is such a system with law μ, we denote its entropy by $\mathrm{H}(X) = \mathrm{H}(\mu)$. Of course, entropy is in fact a (deterministic) function of the law μ of X and not of the (random) values of X.

Intricacies are functionals over such systems (more precisely: over their laws) formalizing and generalizing the neural complexity (1.1) of Edelman-Sporns-Tononi [22]:

Definition 1.1. *A* **system of coefficients** *is a family of numbers*

$$c := (c_S^I : \, I \subset\subset \mathbb{N}^*, \, S \subset I)$$

satisfying, for all I and all $S \subset I$:

$$c_S^I \geq 0, \quad \sum_{S \subset I} c_S^I = 1, \quad \text{and } c_{S^c}^I = c_S^I \tag{1.2}$$

where $S^c := I \setminus S$. *We denote the set of such systems by* $\mathcal{C}(\mathbb{N}^*)$.

The corresponding **mutual information functional** *is* $\mathcal{I}^c : \mathcal{X} \to \mathbb{R}$ *defined by:*

$$\mathcal{I}^c(X) := \sum_{S \subset I} c_S^I \, \mathrm{MI}\,(X_S, X_{S^c})\,.$$

By convention, $\mathrm{MI}\,(X_\emptyset, X_I) = \mathrm{MI}\,(X_I, X_\emptyset) = 0$. *If* $X \in \mathcal{X}(d, I)$ *has law* μ, *we denote* $\mathcal{I}^c(X) = \mathcal{I}^c(\mu)$. $\mathcal{I}^c$ *is* **non-null** *if some coefficient* c_S^I *with* $S \notin \{\emptyset, I\}$ *is not zero.*

An **intricacy** *is a mutual information functional satisfying:*

1. **exchangeability** *(invariance by permutations): if* $I, J \subset\subset \mathbb{N}^*$ *and* $\phi : I \to J$ *is a bijection, then* $\mathcal{I}^c(X) = \mathcal{I}^c(Y)$ *for any* $X := (X_i)_{i \in I}$, $Y := (X_{\phi^{-1}(j)})_{j \in J}$;
2. **weak additivity**: $\mathcal{I}^c(X, Y) = \mathcal{I}^c(X) + \mathcal{I}^c(Y)$ *for any two* independent *systems* $(X_i)_{i \in I}, (Y_j)_{j \in J}$.

Clearly, by (1.1), neural complexity is a mutual information functional with $c_S^I = \frac{1}{|I|+1} \frac{1}{\binom{|I|}{|S|}}$, satisfying exchangeability. Weak additivity is less trivial and will be deduced in Theorem 1.2 below. We remark that the factor $(|I| + 1)$ in the denominator is not present in the original definition in [22] but is necessary for weak additivity and the normalization (1.2) to hold.

1.4. **Main results.** Our first result is a characterization of systems of coefficients c generating an intricacy, i.e. an exchangeable and weak additive mutual information functional. These properties are equivalent to a probabilistic representation of c.

We say that a probability measure λ on $[0, 1]$ is **symmetric** if $\int_{[0,1]} f(x)\, \lambda(dx) = \int_{[0,1]} f(1-x) \lambda(dx)$ for all measurable and bounded functions f.

Theorem 1.2. *Let* $c \in \mathcal{C}(\mathbb{N}^*)$ *be a system of coefficients and* $\mathcal{I}^c$ *the associated mutual information functional.*

1. $\mathcal{I}^c$ *is an intricacy, i.e. exchangeable and weakly additive, if and only if there exists a symmetric probability measure* λ_c *on* $[0, 1]$ *such that*

$$c_S^I = \int_{[0,1]} x^{|S|} (1-x)^{|I|-|S|} \, \lambda_c(dx), \qquad \forall\, S \subseteq I. \tag{1.3}$$

   In this case, if $\{W_c, Y_i, i \in \mathbb{N}^*\}$ *is an independent family such that* W_c *has law* λ_c *and* Y_i *is uniform on* $[0, 1]$, *then*

$$c_S^I = \mathbb{P}(\mathcal{Z} \cap I = S), \qquad \forall\, I \subset\subset \mathbb{N}^*,\ \forall\, S \subset I,$$

   where $\mathcal{Z}$ *is the random subset of* $\mathbb{N}^*$

$$\mathcal{Z} := \{i \in \mathbb{N}^* : Y_i \geq W_c\}.$$

2. λ_c *is uniquely determined by* $\mathcal{I}^c$. *Moreover* $\mathcal{I}^c$ *is non-null iff* $\lambda_c(]0, 1[) > 0$ *and in this case* $c_S^I > 0$ *for all coefficients with* $S \subset I$, $S \notin \{\emptyset, I\}$.

(3) *For the neural complexity* (1.1), *we have*

$$\frac{1}{|I|+1}\frac{1}{\binom{|I|}{|S|}} = \int_{[0,1]} x^{|S|}(1-x)^{|I|-|S|}\,dx, \qquad \forall\ S \subseteq I,$$

i.e., λ_c *in this case is the Lebesgue measure on* $[0,1]$ *and the neural complexity is indeed exchangeable and weakly additive, i.e. an intricacy.*

We discuss other explicit examples in section 2 below.

Our next result concerns the maximal value of intricacies. As discussed above, this is a subtle issue since large intricacy values require compromises. This can also be seen in that intricacies are differences between entropies, see (2.2) and therefore not concave.

The weak additivity of intricacies is the key to how they grow with the size of the system. This property of neural complexity having been brought to the fore, we obtain linear growth and convergence of the growth speed quite easily. The same holds subject to an entropy condition, independently of the softness of the constraint (measured below by the speed at which δ_N converges to 0).

Denote by $\mathcal{I}^c(d,N)$ and $\mathcal{I}^c(d,N,x)$, $x \in [0,1]$, the supremum of $\mathcal{I}^c(X)$ over all $X \in \mathcal{X}(d,N)$, respectively over all $X \in \mathcal{X}(d,N)$ such that $\mathrm{H}(x) = xN\log d$:

$$\mathcal{I}^c(d,N) := \sup\{\mathcal{I}^c(\mu) : \mu \in \mathcal{M}(d,N)\}, \tag{1.4}$$

$$\mathcal{I}^c(d,N,x) := \sup\{\mathcal{I}^c(\mu) : \mu \in \mathcal{M}(d,N),\ \mathrm{H}(\mu) = xN\log d\}. \tag{1.5}$$

Notice that if $x = 0$ or $x = 1$, then $\mathcal{I}^c(d,N,x) = 0$, since this corresponds to, respectively, deterministic or independent systems, for which all mutual information functionals vanish.

Theorem 1.3. *Let* $\mathcal{I}^c$ *be a non-null intricacy and let* $d \geq 2$ *be some integer.*

(1) *The following limits exist for all* $x \in [0,1]$

$$\mathcal{I}^c(d) := \lim_{n\to\infty}\frac{\mathcal{I}^c(d,N)}{N}, \qquad \mathcal{I}^c(d,x) := \lim_{n\to\infty}\frac{\mathcal{I}^c(d,N,x)}{N}, \tag{1.6}$$

and we have the bounds

$$[x \wedge (1-x)]\,\kappa_c \leq \frac{\mathcal{I}^c(d,x)}{\log d} \leq \frac{\mathcal{I}^c(d)}{\log d} \leq \frac{1}{2}, \tag{1.7}$$

where

$$\kappa_c := 2\int_{[0,1]} y(1-y)\,\lambda_c(dy) > 0, \tag{1.8}$$

and λ_c *is defined in Theorem 1.2.*

(2) *Let* $(\delta_N)_{N\geq 1}$ *be any sequence of non-negative numbers converging to zero and* $x \in [0,1]$. *Then*

$$\mathcal{I}^c(d,x) = \lim_{N\to\infty}\frac{1}{N}\sup\left\{\mathcal{I}^c(X) : X \in \mathcal{X}(d,N),\ \left|\frac{\mathrm{H}(X)}{N\log d} - x\right| \leq \delta_N\right\}.$$

Remark 1.4.

1. By considering a set of independent, identically distributed (i.i.d. for short) random variables on $\{0,\dots,d-1\}$, it is easy to see that for any $0 \le h \le N \log d$, there is $X \in \mathcal{X}(d,N)$ such that $\mathrm{H}(X) = h$ and $\mathcal{I}^c(X) = 0$. Hence minimization of intricacies is a trivial problem also under fixed entropy.

2. It follows that for any (x,y), $0 \le x \le 1$ such that $0 \le y < \mathcal{I}^c(d,x)/\log d$, for any N large enough, there exists $X \in \mathcal{X}(d,N)$ with $\mathrm{H}(X) = xN\log d$ and $\mathcal{I}^c(X) = yN\log d$. Observe, for instance, that $\mathcal{I}^c$ is continuous on the contractile space $\mathcal{M}(d,N)$.

3. In the above theorem, the assumption that each variable X_i takes values in a set of cardinality d can be relaxed to $\mathrm{H}(X_i) \le \log d$. It can be shown that this does not change $\mathcal{I}^c(d)$ or $\mathcal{I}^c(d,x)$.

Thus maximal intricacy grows linearly in the size of the system. What happens if we restrict to smaller classes of systems, enjoying particular symmetries? Since intricacies are exchangeable, their value does not change if we permute the variables of a system. Therefore it is particularly natural to consider (finite) exchangeable families.

We denote by $\mathrm{EX}(d,N)$ the set of random variables $X \in \mathcal{X}(d,N)$ which are exchangeable, i.e., for all permutations σ of $\{1,\dots,N\}$, $X := (X_1,\dots,X_N)$ and $X_\sigma := (X_{\sigma(1)},\dots,X_{\sigma(N)})$ have the same law.

Theorem 1.5. *Let $\mathcal{I}^c$ be an intricacy.*

(1) Exchangeable systems have small intricacies. More precisely

$$\sup_{X\in \mathrm{EX}(d,N)} \mathcal{I}^c(X) = o(N^{2/3+\epsilon}), \qquad N \to +\infty,$$

for any $\epsilon > 0$. In particular

$$\lim_{N\to\infty} \frac{1}{N} \max_{X\in \mathrm{EX}(d,N)} \mathcal{I}^c(X) = 0.$$

(2) For N large enough and fixed d, maximizers of $\mathcal{X}(d,N) \ni X \mapsto \mathcal{I}^c(X)$ are neither unique nor exchangeable.

By the first assertion, exchangeability of the intricacies is not inherited by their maximizers. Indeed, exchangeable systems are very far from maximizing, since the maximum of $\mathcal{I}^c$ over $\mathrm{EX}(d,N)$ is $o(N^p)$ for any $p > 2/3$ whereas the maximum of $\mathcal{I}^c$ over $\mathcal{X}(d,N)$ is proportional to N. This "spontaneous symmetry breaking" again suggests the complexity of the maximizers. We remark that numerical estimates suggest that the intricacy of any $X \in \mathrm{EX}(d,N)$ is in fact bounded by $\mathrm{const}\log N$.

The second assertion of Theorem 1.5 follows from the first one: for N sufficiently large, the maximal intricacy is not attained at an exchangeable law; therefore, by permuting a system with maximal intricacy we obtain different laws, all with the same maximal intricacy.

We finally turn to a property of exact maximizers, namely that their support is concentrated on a small subset of all possible configuration:

Theorem 1.6. *Let $\mathcal{I}^c$ be a non-null intricacy. let $d \geq 2$. For N a large enough integer, the following holds. For any X maximizing $\mathcal{I}^c$ over $\mathcal{X}(d,N)$, law μ of X has small support, i.e.*

$$\#\{\omega \in \Lambda_{d,N} : \mu(\{\omega\}) = 0\} \geq \mathrm{const}\, d^N$$

for some const > 0.

1.5. **Further questions.** As noted above, the exact computation of the functions $\mathcal{I}^c(d)$ and $\mathcal{I}^c(d,x)$ from Theorem 1.3 in terms of their probabilistic representation from Theorem 1.2 will be the subject of [5] where we shall study systems with intricacy close to the maximum.

Second, to apply intricacy one needs to compute it for systems of interests. It might be possible to compute it exactly for some simple physical systems, like the Ising model. A more ambitious goal would be to consider more complex models, like spin glasses, to analyze the possible relation between intricacy and frustration [21].

A more general approach would be to get rigorous estimates from numerical ones (see [22] for some rough computations). A naive approach results in an exponential complexity and thus begs the question of more efficient algorithms, perhaps probabilistic ones. A related question is the design of statistical estimators for intricacies. These estimators should be able to decide many-variables correlations, which might require *a priori* assumptions on the systems.

Third, one would to understand the intricacy from a dynamical point of view: which physically reasonable processes (say with dynamics defined in terms of local rules) can lead to high intricacy systems and at what speeds?

Fourthly, one could consider the natural generalization of intricacies, already proposed in [22] but not explored further, is given in terms of general partitions π of I: if $\pi = \{S_1, ..., S_k\}$ with $\cup_i S_i = I$ and $S_i \cap S_j = \emptyset$ for $i \neq j$, then we can set

$$\mathrm{MI}(X_\pi) := \mathrm{H}(X_{S_1}) + \cdots + \mathrm{H}(X_{S_k}) - \mathrm{H}(X), \qquad X \in \mathcal{X}(d,I), \tag{1.9}$$

and for some non-negative coefficients $(c_\pi)_\pi$

$$\mathcal{J}^c(X) := \sum_\pi c_\pi \,\mathrm{MI}(X_\pi). \tag{1.10}$$

Most results of this paper extend to the case where the coefficients $(c_\pi)_\pi$ have a probabilistic representation in terms of the so-called Kingman paintbox construction [4, §2.3], see Remark 3.4 below.

One might also be interested to extend the definition of intricacy to infinite (e.g., stationary) processes, continuous or structured systems, e.g., taking into account a connectivity or dependence graph (such constraints have been considered in numerical experiments performed by several authors [2, 8, 18]).

Finally, our work leaves out the properties of exact maximizers for a given size. As of now, we have no description of them except in very special cases (see Examples 2.9 and 2.10 below) and we do not know how many there are, or even if they are always in finite number. We do not have reasonably efficient ways to determine the maximizers which we expect to lack a simple description in light of the lack of symmetry established in Theorem 1.5.

1.6. **Organization of the paper.** In Sec. 2, we discuss the definition of intricacies, giving some basic properties and examples. Sec. 3 proves Theorem 1.2, translating the weak additivity of an intricacy into a property of its coefficients. As a by-product, we obtain a probabilistic representation of all intricacies. We check that neural complexity corresponds to the uniform law on $[0,1]$. In Sec. 4 we prove Theorem 1.3 by showing the existence of the limits $\mathcal{I}^c(d)$, $\mathcal{I}^c(d,x)$. Finally, in Sec. 5 we prove Theorem 1.5 and, in Sec. 6, Theorem 1.6. An Appendix recalls some basic facts from information theory for the convenience of the reader and to fix notations.

2. Intricacies

2.1. **Definition.** We begin by a discussion of the definition 1.1 above of intricacies. As $\mathrm{MI}(X_S, X_{S^c}) = \mathrm{MI}(X_{S^c}, X_S)$, the symmetry condition $c^I_{S^c} = c^I_S$ can always be satisfied by replacing c^I_S with $\frac{1}{2}(c^I_S + c^I_{S^c})$ without changing the functional. Also $\sum_{S\subset I} c^I_S = 1$ is simply an irrelevant normalization when studying systems with a given index set I.

The following mutual information functionals will be proved to be intricacies in section 3.

Definition 2.1. *The intricacy $\mathcal{I}$ of* **Edelman-Sporns-Tononi** *is defined by its coefficients:*

$$c^I_S = \frac{1}{|I|+1}\frac{1}{\binom{|I|}{|S|}}. \tag{2.1}$$

*For $0 < p < 1$, the p-***symmetric intricacy** *$\mathcal{I}^p(X)$ is:*

$$c^I_S = \frac{1}{2}\left(p^{|S|}(1-p)^{|I\setminus S|} + (1-p)^{|S|}p^{|I\setminus S|}\right).$$

For $p = 1/2$, this is the **uniform intricacy** *$\mathcal{I}^U(X)$ with:*

$$c^I_S = 2^{-|I|}.$$

It is not obvious that the three above mutual information functionals are weakly additive, but this will follow easily from Lemma 3.7 below. Proposition 3.5 below describes all intricacies.

Remark 2.2. The coefficients of the Edelman-Sporns-Tononi intricacy $\mathcal{I}$ ensure that subsystems of all sizes contribute significantly to the intricacy. This is in sharp contrast to the p-symmetric coefficients for which subsystems of size far from pN or $(1-p)N$ give a vanishing contribution when N gets large.

Remark 2.3. The global $1/(|I|+1)$ factor in $\mathcal{I}$ is not present in [22], which did not compare systems of different sizes. However it is required for weak additivity.

2.2. **Basic Properties.** We prove some general and easy properties of intricacies. Recall that $\mathcal{X}(d,N)$ is the set of $\Lambda_{d,N}$-valued random variables, where $\Lambda_{d,N} = \{0,\dots,d-1\}^N$. We identify it with the standard simplex in $\mathbb{R}^{d^N}$ in the obvious way.

Lemma 2.4. *Let $\mathcal{I}^c$ be a mutual information functional. For each $d \geq 2$ and $N \geq 1$, $\mathcal{I}^c : \mathcal{M}(d,N) \to \mathbf{R}$ is continuous. In particular, the suprema $\mathcal{I}^c(d,N)$ and $\mathcal{I}^c(d,N,x)$, introduced in* (1.4) *and* (1.5)*, are achieved.*

If $\mathcal{I}^c$ is a non-null intricacy, then it is neither convex nor concave.

Proof. Continuity is obvious and existence of the maximum follows from the compactness of the finite-dimensional simplex $\mathcal{M}(d,N)$. To disprove convexity and concavity of non-null intricacies, we use the following examples. Pick I with at least two elements, say 1 and 2. Observe that $K := c^I_{\{1\}} + c^I_{\{2\}}$ is positive by the non-degeneracy of $\mathcal{I}^c$ (see Lemma 3.8 below). Fix $d \geq 2$.

First, for $i = 0,1$, let μ_i over $\{0,\dots,d-1\}^I$ be defined by $\mu_i(i,i,0,\dots,0) = 1$. We have:

$$\mathcal{I}^c\left(\frac{\mu_0+\mu_1}{2}\right) \geq K \cdot \log d > \frac{\mathcal{I}^c(\mu_0)+\mathcal{I}^c(\mu_1)}{2} = 0.$$

Second, let ν_0 be defined by $\nu_0(0,0,0,\dots,0) = \nu_0(1,1,0,\dots,0) = 1/2$ and ν_1 by $\nu_1(0,1,0,\dots,0) = \nu_1(1,0,0,\dots,0) = 1/2$. We have:

$$\mathcal{I}^c\left(\frac{\nu_0+\nu_1}{2}\right) = 0 < K \cdot \log d \leq \frac{\mathcal{I}^c(\nu_0)+\mathcal{I}^c(\nu_1)}{2}.$$

□

The following expression of an intricacy as a non-convex combination of the entropy of subsystems is crucial to its understanding.

Lemma 2.5. *For any intricacy $\mathcal{I}^c$ and $X \in \mathcal{X}(d,N)$*

$$\mathcal{I}^c(X) = 2\left(\sum_{S\subset I} c^I_S \,\mathrm{H}(X_S)\right) - \mathrm{H}(X). \tag{2.2}$$

Proof. The result readily follows from: $\mathrm{MI}(X,Y) = \mathrm{H}(X) + \mathrm{H}(Y) - \mathrm{H}(X,Y)$, $c^I_S = c^I_{S^c}$, and $\sum_S c^I_S = 1$. □

We introduce the notation

$$\mathrm{MI}(S) := \mathrm{MI}(X_S, X_{I\setminus S})$$

which will be used only when the understood dependence on X and I is clear.

Lemma 2.6. *For any intricacy $\mathcal{I}^c$ and any system $X \in \mathcal{X}(d,N)$*

$$0 \leq \mathcal{I}^c(X) \leq \frac{N}{2}\log d.$$

Proof. The inequalities follow from basic properties of the mutual information (see the Appendix):

$$0 \le \mathrm{MI}(S) \le \min\{\mathrm{H}(X_S), \mathrm{H}(X_{S^c})\} \le \min\{|S|, N-|S|\} \log d \le \frac{N}{2} \log d.$$

□

2.3. **Simple examples.** We give some examples of finite systems and compute their intricacies both for illustrative purposes and for their use in some proofs below.

Let X_i take values in $\{0, \dots, d-1\}$ for all $i \in I$, a finite subset of $\mathbb{N}^*$. The first two examples show that total order and total disorder make the intricacy vanish.

Example 2.7 (Total disorder)**.** If the variables X_i are independent then each mutual information is zero and therefore: $\mathcal{I}^c(X) = 0$. □

Example 2.8 (Total order)**.** If each X_i is a.s. equal to a constant c_i in $\{0, \dots, d-1\}$, then, for any $S \neq \emptyset$, $\mathrm{H}(X_S) = 0$. Hence, $\mathcal{I}^c(X) = 0$. □

For $N = 2, 3$, each mutual information can be maximized separately: there is no frustration and it is easy to determine the maximizers of non-null intricacies.

Example 2.9 (Case $N = 2$)**.** Let first $N = 2$ and $\mathcal{I}^c$ be a non-null intricacy. Then by Theorem 1.2 $c_S^I = c_{|S|}^{|I|}$ and therefore

$$\mathcal{I}^c(X) = \left(c_{\{1\}}^{\{1,2\}} + c_{\{2\}}^{\{1,2\}}\right) \mathrm{MI}(X_1, X_2) = 2c_1^2\, \mathrm{MI}(X_1, X_2), \qquad X \in \mathcal{X}(d, 2),$$

and moreover $c_1^2 > 0$. Therefore the maximizers of $\mathcal{I}^c$ over $\mathcal{X}(d, 2)$ are the maximizers of $X \mapsto \mathrm{MI}(X_1, X_2)$. By the discussion in subsection A.3 of the appendix, we have that $\mathrm{MI}(X_1, X_2) \le \min\{\mathrm{H}(X_1), \mathrm{H}(X_2)\}$. Now, $\mathrm{MI}(X, Y) = \mathrm{H}(X_1) = \mathrm{H}(X_2)$ iff each variable is a function of the other.

Therefore, the maximizers are exactly the following systems $X = (X_1, X_2)$. X_1 is a uniform r.v. over $\{0, \dots, d-1\}$ and the other is a deterministic function of the first. $X_2 = \sigma(X_1)$ for a given permutation σ of $\{0, \dots, d-1\}$. In the case of the neural complexity, $\max_{X \in \mathcal{X}(d,2)} \mathcal{I}(X) = (\log d)/3$. □

Example 2.10 (Case $N = 3$)**.** Let $N = 3$ and $I := \{1, 2, 3\}$. By Theorem 1.2, $c_S^I = c_{|S|}^{|I|}$, $c_1^3 = c_2^3$ and therefore

$$\mathcal{I}^c(X) = 2c_1^3 \left(\mathrm{MI}(X_1, X_{\{2,3\}}) + \mathrm{MI}(X_2, X_{\{1,3\}}) + \mathrm{MI}(X_3, X_{\{1,2\}})\right),$$

and moreover $c_1^3 > 0$. Here we simultaneously maximize each of these mutual informations. The optimal choice is a system (X_1, X_2, X_3) where every pair (X_i, X_j), $i \neq j$, is uniform over $\{0, \dots, d-1\}^2$, and the third variable is a function of (X_i, X_j). This is realized iff (X_1, X_2) is uniform over $\{0, \dots, d-1\}^2$ and $X_3 = \phi(X_1, X_2)$, where ϕ is a (deterministic) map such that, for any $i \in \{0, \dots, d-1\}$, $\phi(i, \cdot)$ and $\phi(\cdot, i)$ are permutations of $\{0, \dots, d-1\}$. For instance: $\phi(x_1, x_2) = x_1 + x_2 \mod d$. In the case of the neural complexity, $\max_{X \in \mathcal{X}(d,3)} \mathcal{I}(X) = (\log d)/2$. □

The maximizers of examples 2.9 and 2.10 are very special. For instance, they are exchangeable, contrarily to the case of large N according to Theorem 1.5. For $N = 4$ and beyond it is no longer possible to separately maximize each mutual information and we do not have an explicit description of the maximizers. We shall however see that, as in the above examples, maximizers have small support, see Proposition 1.6.

Remark 2.11. Example 2.10 has an interesting interpretation: for $N = 3$, a system with large intricacy shows in a simple way a combination of *differentiation* and *integration*, as it is expected in the biological literature, see the Introduction. Indeed, any subsystem of two variables is independent (differentiation), while the whole system is correlated (integration).

Another interesting case is that of a large system where one variable is free and all others follow it deterministically.

Example 2.12 (A totally synchronized system)**.** Let X_1 be a uniform $\{0,\ldots,d-1\}$-valued random variable. We define now $(X_2,\ldots,X_N) := \phi(X_1)$, where ϕ is any deterministic map from $\{0,\ldots,d-1\}$ to $\{0,\ldots,d-1\}^{N-1}$. Then, for any $S \neq \emptyset$, $\mathrm{H}(X_S) = \log d$ and, if additionally $S^c \neq \emptyset$, $\mathrm{H}(X_S|X_{S^c}) = 0$ so that each mutual information $\mathrm{MI}(X_S, X_{S^c})$ is $\log d$ if $S \notin \{\emptyset, I\}$. Hence,

$$\mathcal{I}^c(X) = \sum_{S\subset I\setminus\{\emptyset,I\}} c_S^I \cdot \log d = \left(1 - c_\emptyset^I - c_I^I\right) \log d. \qquad \square$$

In the next example we build for every $x \in]0,1[$ a system $X \in \mathcal{X}(d,2)$ with entropy $\mathrm{H}(X) = x \log d^2$ and positive intricacy.

Example 2.13 (A system with positive intricacy and arbitrary entropy)**.** Let first $x \in]0,1/2]$. Let X_1 be $\{0,\ldots,d-1\}$-valued with $\mathrm{H}(X_1) = 2x \log d$. Such a variable exists because entropy is continuous over the connected simplex of probability measures on $\{0,\ldots,d-1\}$ and attains the values 0 over a Dirac mass and $\log d$ over the uniform distribution. We define now $X_2 := X_1$ and $X := (X_1, X_2) \in \mathcal{X}(d,2)$. Therefore $\mathrm{H}(X) = 2x \log d = x \log d^2$, $\mathrm{MI}(X_1, X_2) = \mathrm{H}(X_1) = 2x \log d$ and, arguing as in Lemma 2.9

$$\mathcal{I}^c(X) = 2c_1^2 \ \mathrm{MI}(X_1, X_2) = 4x \, c_1^2 \, \log d > 0.$$

Let now $x \in]1/2,1[$. Let (Y_1, Y_2, B) be an independent triple such that Y_i is uniform over $\{0,\ldots,d-1\}$ and B is Bernoulli with parameter $p \in [0,1]$ and set

$$X_1 := Y_1, \qquad X_2 := \mathbb{1}_{(B=0)}\, Y_1 + \mathbb{1}_{(B=1)}\, Y_2, \qquad X := (X_1, X_2).$$

Then both X_1 and X_2 are uniform on $\{0,\ldots,d-1\}$. On the other hand, it is easy to see that $\mathrm{H}(X)$, as a function of $p \in [0,1]$, interpolates continuously between $\log d$ and $2\log d$. Thus, for every $x \in]1/2,1[$ there is a $p \in [0,1]$ such that $\mathrm{H}(X) = x \log d^2$. In this case $\mathrm{MI}(X_1, X_2) = 2(1-x) \log d$ and we obtain

$$\mathcal{I}^c(X) = 2c_1^2 \ \mathrm{MI}(X_1, X_2) = 4(1-x)\, c_1^2 \, \log d > 0. \qquad \square$$

Intricacy can indeed reach over $\mathcal{X}(d,N)$ the order N of Lemma 2.6, as the next example shows.

Example 2.14 (Systems with uniform intricacy proportional to N)**.** Let us fix $d \geq 2$. For $N \geq 2$, we are going to build a system $(X_i)_{i\in I}$, $I = \{1,\dots,N\}$, over the alphabet $\{0,\dots,d^2-1\}$ for which $\mathcal{I}^U(X)/N$ converges to $(\log d^2)/4$; later, in Example 3.10, we shall generalize this to an arbitrary intricacy.

Let $Y_1,\dots,Y_N$ be i.i.d. uniform $\{0,\dots,d-1\}$-valued random variables and define $X_i := Y_i + dY_{i+1}$ for $i = 1,\dots,N-1$, $X_N := Y_N$. Note that $X \in \mathcal{X}(d^2,N)$ and $\mathrm{H}(X) = N\log d = (N/2)\log d^2$. For $S \subset I$, set

$$\Delta_S := \{k = 1,\dots,N-1 : \mathbb{1}_S(k) \neq \mathbb{1}_S(k+1)\},$$
$$U_S := \{k = 1,\dots,N-1 : \mathbb{1}_S(k) = 1 \neq \mathbb{1}_S(k+1)\}.$$

Observe that $\mathrm{H}(X_S) = (|S| + |U_S|)\log d$. Indeed, this is given by $\log d$ times the minimal number of Y_i needed to define X_S; every $k \in S$ counts for one if $k \in S\setminus U_S$, for two if $k \in U_S$; therefore we find $|S| - |U_S| + 2|U_S| = |S| + |U_S|$. Moreover, $|U_S| + |U_{S^c}| = |\Delta_S|$. Therefore

$$\mathrm{MI}(S) = (|U_S| + |S| + |U_{S^c}| + |S^c| - N)\log d = |\Delta_S|\,\log d.$$

Moreover we have a bijection:

$$S \in \{0,1\}^{\{1,\dots,N\}} \mapsto (\mathbb{1}_S(1), \Delta_S) \in \{0,1\} \times \{0,1\}^{\{1,\dots,N-1\}}.$$

Hence:

$$\begin{aligned}\frac{\mathcal{I}^U(X)}{\log d} &= 2^{-N}\sum_{S\subset I}|\Delta_S| = 2^{-N}\times 2\sum_{\Delta\subset\{1,\dots,N-1\}}|\Delta| = 2^{-N+1}\sum_{k=0}^{N-1}\binom{N-1}{k}k \\ &= 2^{-N+1}(N-1)2^{N-2} = \frac{N-1}{2}.\end{aligned}$$

Therefore for this $X \in \mathcal{X}(d^2,N)$:

$$\mathcal{I}^U(X) = \frac{N-1}{4}\log(d^2). \qquad \square$$

The following example will be useful to show that an intricacy $\mathcal{I}^c$ determines its coefficients $c \in \mathcal{C}(\mathbb{N}^*)$ in Lemma 3.2 below.

Example 2.15 (A system with a synchronized sub-system)**.** We consider a system of uniform variables, with a subset of equal ones and the remainder independent. More precisely, let $I \subset\subset \mathbb{N}^*$, $\emptyset \neq K \subset I$ and fix $i_0 \in K$. $(X_i)_{i\in I} \in \mathcal{X}(d,I)$ is the system satisfying:

(i) the family $X_{K^c\cup\{i_0\}}$ is uniform on $\{0,\dots,d-1\}^{K^c\cup\{i_0\}}$;
(ii) $X_i = X_{i_0}$ for all $i \in K$.

It follows that

$$\mathrm{H}(X_S) = \big(|S\setminus K| + \mathbb{1}_{(S\cap K\neq\emptyset)}\big)\log d$$

and therefore

$$\mathrm{MI}(S) = \big(\mathbb{1}_{(S\cap K\neq\emptyset)} + \mathbb{1}_{(S^c\cap K\neq\emptyset)} - 1\big)\log d,$$

i.e. $\mathrm{MI}(S) = 0$ unless S and S^c both intersect K and then $\mathrm{MI}(S) = \log d$. Thus

$$\mathcal{I}^c(X) = \log d\sum_{S\subset I}c_S^I\,\mathbb{1}_{(\emptyset\neq S\cap K\neq K)}, \qquad \mathrm{H}(X) = (|K^c|+1)\log d. \qquad \square$$

3. Weak additivity, projectivity and representation

In this section we prove Theorem 1.2, by studying the additivity of mutual information functionals and characterizing it in terms of the coefficients. We establish a probabilistic representation of all intricacies and check that the neural complexity is indeed an intricacy. We conclude this section by some useful consequences of this representation.

Throughout this section, $X = (X_i)_{i\in I}$ and $Y = (Y_i)_{i\in J}$, will be two systems defined on the same probability space and we shall consider the joint family $(X, Y) = \{X_i, Y_j : i \in I, j \in J\}$. (X, Y) is again a system and its index set is the disjoint union $I \sqcup J$ of I and J.

3.1. **Projectivity and Additivity.** We show that weak additivity and exchangeability can be read off the coefficients and that non-null intricacies are neither sub-additive nor super-additive.

Proposition 3.1. *Let $\mathcal{I}^c$ be a mutual information functional. Then*

(1) $\mathcal{I}^c$ is exchangeable if and only if c^I_S depends only on $|I|$ and $|S|$

(2) $\mathcal{I}^c$ is weakly additive if and only if the coefficients are **projective***, i.e., satisfy*

$$\forall\, I \subset\subset \mathbb{N}^*,\ \forall\, J \subset\subset \mathbb{N}^* \setminus I,\ \forall\, S \subset I, \qquad c^I_S = \sum_{T\subset J} c^{I\sqcup J}_{S\sqcup T}. \tag{3.1}$$

(3) Let $\mathcal{I}^c$ be an intricacy. Then, for non-necessarily independent systems X, Y, we have: $\mathcal{I}^c(X, Y) \geq \max\{\mathcal{I}^c(X), \mathcal{I}^c(Y)\}$ and the approximate additivity:

$$|\mathcal{I}^c(X) + \mathcal{I}^c(Y) - \mathcal{I}^c(X, Y)| \leq \mathrm{MI}(X, Y);$$

(4) $\mathcal{I}^c$ can fail to be super-additive or sub-additive.

To prove this proposition we shall need the following fact:

Lemma 3.2. *Let $d \geq 2$ and I be a finite set. The data $\mathcal{I}^c(X)$ for $X \in \mathcal{X}(d, J)$ for all $J \subset\subset I$ determine $c \in \mathcal{C}(I)$.*

Proof. Using $c^I_{S^c} = c^I_S$, we restrict ourselves to coefficients with $|S| \leq |S^c|$, i.e., $|S| \leq |I|/2$. Let us first consider a system $(X_i)_{i\in I} \in \mathcal{X}(d, I)$ where all variables are equal: $X_i = X_j$ for all $i, j \in I$ and X_i is uniform on $\{0, \ldots, d-1\}$. Then $\mathrm{MI}(S) := \mathrm{MI}(X_S, X_{S^c}) = 0$ for $S = \emptyset$ or $S = I$, otherwise $\mathrm{MI}(S) = \log d$. Hence, using the normalization $1 = \sum_S c^I_S$:

$$1 - \frac{\mathcal{I}^c(X)}{\log d} = \sum_S c^I_S - \sum_{\emptyset \subsetneq S \subsetneq I} c^I_S = c^I_\emptyset + c^I_I.$$

In particular, $c^I_\emptyset = c^I_I = (1 - \mathcal{I}^c(X)/\log d)/2$.

For each $K \subset I$, let X^K be the system as in Example 2.15. Fix $i_0 \in K$. Recall that $\mathrm{MI}(S) := \mathrm{MI}(X_S, X_{S^c})$ is 0 if $S \supset K$ or $S^c \supset K$, and is $\log d$ otherwise. Assume by induction that, for $1 \leq s \leq |I|/2$, c^I_S is determined for $|S| < s$ (a trivial assertion for $s = 1$). Picking $K \subset I$ with $|K| = |I| - s \geq |I|/2 \geq |K^c| = s$, we get:

- if $|S| < s$, we say nothing of $\mathrm{MI}(S)$ but will use the inductive assumption;
- if $S = K$ or $S = K^c$, then $\mathrm{MI}(S) = 0$;

- if $s \le |S| \le |K|$, $S \supset K$ implies $S = K$, $S \subset K^c$ implies $S = K^c$ since $s = |K^c|$. In all other cases: $\mathrm{MI}(S) = \log d$.

Therefore,

$$\begin{aligned}
\frac{\mathcal{I}^c(X^K)}{\log d} &= 2\sum_{S\subset I} c_S^I \frac{\mathrm{MI}(S)}{\log d} - \frac{\mathrm{H}(X^K)}{\log d} \\
&= 4 \sum_{|S|<|I|/2} c_S^I \frac{\mathrm{MI}(S)}{\log d} + 2 \sum_{|S|=|I|/2} c_S^I \frac{\mathrm{MI}(S)}{\log d} - \frac{\mathrm{H}(X^K)}{\log d} \\
&= 4\sum_{|S|<s} c_S^I \frac{\mathrm{MI}(S)}{\log d} + 4 \sum_{s\le |S|<|I|/2} c_S^I + 2 \sum_{|S|=|I|/2} c_S^I \; - \; 2(c_K^I + c_{K^c}^I) - \frac{\mathrm{H}(X^K)}{\log d}
\end{aligned}$$

(the sum over $|S| = |I|/2$ is non-zero only if $|I|$ is even). Using $\sum_S c_S^I = 1$ and $c_S^I = c_{S^c}^I$, we get:

$$\frac{\mathcal{I}^c(X^K)}{\log d} + \frac{\mathrm{H}(X)}{\log d} - 2 = 2\sum_{S\subset I} c_S^I \left(\frac{\mathrm{MI}(S)}{\log d} - 1\right) = 4 \sum_{|S|<s} c_S^I \left(\frac{\mathrm{MI}(S)}{\log d} - 1\right) - 4c_{K^c}^I.$$

It follows that $c_K^I = c_{K^c}^I$ is determined for any K with $|K| = s$. This completes the induction step and the proof of the lemma. $\square$

Proof of Proposition 3.1. The characterization of exchangeability is a direct consequence of Lemma 3.2.

Let us prove the second point. We first check that weak additivity implies projectivity. For any $X \in \mathcal{X}(d, I)$ with $I \subset\subset \mathbb{N}^*$ and $J \subset\subset \mathbb{N}^* \setminus I$, we have:

$$\mathcal{I}^c(X) = \mathcal{I}^c(X, Z) = \sum_{S\subset I}\sum_{T\subset J} c_{S\sqcup T}^{I\cup J}\,\mathrm{MI}(X_S, X_{S^c})$$

for $Z = (Z_j)_{j\in J}$ with each Z_j a.s. constant and therefore independent of X. Lemma 3.2 then implies that (3.1) holds. Moreover, (A.6) yields the monotonicity claim of point (2).

For the approximate additivity, we consider (A.7) for any $S \subset I$, $T \subset J$:

$$\mathrm{MI}((X_S, Y_T), (X_{S^c}, Y_{T^c})) = \mathrm{MI}(X_S, X_{S^c}) + \mathrm{MI}(Y_T, Y_{T^c}) \pm \mathrm{MI}(X, Y)$$

where $\pm\,\mathrm{MI}(X, Y)$ denotes a number belonging to $[-\,\mathrm{MI}(X, Y), \mathrm{MI}(X, Y)]$. The projectivity now gives:

$$\begin{aligned}
\mathcal{I}^c(X, Y) &= \sum_{S\subseteq I, T\subseteq J} c_{S\sqcup T}^{I\sqcup J}\,\mathrm{MI}(S \sqcup T) \\
&= \sum_{S\subseteq I, T\subseteq J} c_{S\sqcup T}^{I\sqcup J}(\mathrm{MI}(X_S, X_{S^c}) + \mathrm{MI}(Y_T, Y_{T^c}) \pm \mathrm{MI}(X, Y)) \\
&= \mathcal{I}^c(X) + \mathcal{I}^c(Y) \pm \mathrm{MI}(X, Y).
\end{aligned}$$

This is the approximate additivity of point (2). If X and Y are independent, then $\mathrm{MI}(X, Y) = 0$, proving the weak additivity.

We finally give the counter-examples. For sub-additivity, it is enough to assume the intricacy to be non-null and to consider $X = Y$ a single random variable uniform on $\{1, 2\}$ and compute:

$$\mathcal{I}^c(X) = \mathcal{I}^c(Y) = 0 \text{ whereas } \mathcal{I}^c(X, Y) = 2c_1^2 \log 2 > 0.$$

For super-additivity, we assume $c_\emptyset^I + c_I^I < \frac{1}{2} + \frac{c_\emptyset^{I\sqcup I} + c_{I\sqcup I}^{I\sqcup I}}{2}$ and take $X = Y$ a collection of $N = |I|$ copies of the same variable uniform over $\{0, 1\}$. Then $\mathrm{MI}(S) = \log 2$ except if $S \in \{\emptyset, I\}$, in which case $\mathrm{MI}(S) = 0$. By example 2.12

$$\frac{\mathcal{I}^c(X, Y)}{\log 2} = 1 - c_\emptyset^{I\sqcup I} - c_{I\sqcup I}^{I\sqcup I} < 2\left(1 - c_\emptyset^I - c_I^I\right) = \frac{\mathcal{I}^c(X) + \mathcal{I}^c(Y)}{\log 2}.$$

□

3.2. **Probabilistic representation of intricacies.** In this section, we give the probabilistic representation for intricacies. This will provide us with a way to estimate the maximal value of intricacy for large systems in [5]. For notational convenience, we consider intricacies over the positive integers $\mathbb{N}^*$.

We say that a random variable W over $[0, 1]$ is **symmetric** if W and $1 - W$ have the same law. A measure on $[0, 1]$ is symmetric if it is the law of a symmetric random variable.

Proposition 3.3. *Let $\mathcal{I}^c$ be a mutual information functional defined by some system of coefficients $c \in \mathcal{C}(\mathbb{N}^*)$ over some infinite index set, which we assume to be $\mathbb{N}^*$ for notational convenience.*

(1) $\mathcal{I}^c$ is an intricacy, i.e., it is exchangeable and weakly additive, if and only if there exists a symmetric random variable W_c over $[0, 1]$ with law λ_c such that for all $I \subset\subset \mathbb{N}^*$ and $S \subset I$

$$c_S^I = \mathbb{E}\left(W_c^{|S|}(1 - W_c)^{|I|-|S|}\right) = \int_{[0,1]} x^{|S|}(1 - x)^{|I|-|S|}\, \lambda_c(dx). \tag{3.2}$$

(2) Formula (3.2) is equivalent to

$$c_S^I = \mathbb{P}(\mathcal{Z} \cap I = S), \qquad \forall\, I \subset\subset \mathbb{N}^*,\ \forall\, S \subset I, \tag{3.3}$$

where $\mathcal{Z}$ is the random subset of $\mathbb{N}^*$

$$\mathcal{Z} := \{i \in \mathbb{N}^* : Y_i \geq W_c\}, \tag{3.4}$$

with $(Y_i)_{i\geq 1}$ an i.i.d. sequence of uniform random variables on $[0, 1]$, independent of W_c.

(3) If $\mathcal{I}^c$ is an intricacy, then the law λ_c of W_c is uniquely determined by $\mathcal{I}^c$. Moreover for all $X \in \mathcal{X}(d, I)$ independent of $\mathcal{Z}$

$$\mathcal{I}^c(X) = \mathbb{E}(\mathrm{MI}(\mathcal{Z} \cap I)), \qquad \mathrm{MI}(S) := \mathrm{MI}(X_S, X_{I\setminus S}).$$

Remark 3.4. The definition (3.4) of the random set $\mathcal{Z}$ is a particular case of the so-called *Kingman paintbox* construction, see [4, §2.3]. In this setting, it yields a random exchangeable partition of $\mathbb{N}^*$ into a subset $\mathcal{Z}$ and its complement, each with asymptotic density a.s. equal to W_c, respectively $1 - W_c$. Therefore it is natural to expect a similar probabilistic representation for coefficients $(c_\pi)_\pi$ of exchangeable and weakly additive generalized functionals defined in (1.9) and (1.10).

After the proof of the proposition we give the measures μ, μ^U, μ^p representing respectively $\mathcal{I}$, $\mathcal{I}^U$ and $\mathcal{I}^p$. We start with the following

Lemma 3.5. *Let $\mathcal{C}(\mathbb{N}^*)$ be the set of systems of coefficients of intricacies. Let $\mathcal{PS}([0,1])$ be the set of symmetric probability measures λ on $[0,1]$. Then, the map $\lambda \mapsto c$ defined from $\mathcal{PS}([0,1])$ to $\mathcal{C}(\mathbb{N}^*)$ according to ($n := |I|$, $k := |S|$):*

$$c_S^I = c_k^n = \int_{[0,1]} x^k (1-x)^{n-k}\, \lambda(dx), \qquad \forall S \subset I \subset\subset \mathbb{N}^*, \tag{3.5}$$

is a bijection.

Proof of Lemma 3.5. We first show that for an exchangeable weakly additive $\mathcal{I}^c$, there exists a probability measure λ on $[0,1]$ such that

$$c_n^{n+k} = \int_{[0,1]} x^n (1-x)^k\, \lambda(dx), \qquad n \geq 1,\, k \geq 0 \tag{3.6}$$

i.e., the main claim of the Lemma, up to a convenient renumbering. We need the following classical moment result, see e.g. [9, VII.3].

Lemma 3.6. *Let $(a_n)_{n\geq 1}$ be a sequence of numbers in $[0,1]$. We define $(Da)_n := a_n - a_{n+1}$, $n \geq 1$. There exists a probability measure λ on $[0,1]$ such that $a_n = \int x^n\, \lambda(dx)$ if and only if*

$$(D^k a)_n \geq 0, \qquad \forall\, n \geq 1,\ \forall k \geq 1.$$

Moreover such λ is unique.

Remark that, setting $N = |I|$ and $M = |J|$, projectivity is equivalent to

$$c_k^N = \sum_{\ell=0}^{M} c_{k+\ell}^{M+N} \binom{M}{\ell}, \qquad \forall\, 0 \leq k \leq N. \tag{3.7}$$

For $M = 1$ we obtain

$$c_k^{N+1} + c_{k+1}^{N+1} = c_k^N, \qquad \forall\, 0 \leq k \leq N. \tag{3.8}$$

Let us set $m_n := c_n^n$. One proves easily by (3.8) and recurrence on k that

$$(D^k m)_n = c_n^{n+k} \in [0,1], \qquad \forall\, k, n \geq 1.$$

Therefore $(m_n)_{n\geq 1}$ defines a unique measure λ satisfying (3.6) for $k = 0$. (3.6) for general k follows by induction from:

$$\begin{aligned} c_n^{n+k+1} = c_n^{n+k} - c_{n+1}^{n+k+1} &= \int_{[0,1]} \left[x^n (1-x)^k - x^{n+1}(1-x)^k \right] d\lambda \\ &= \int_{[0,1]} x^n (1-x)^{k+1}\, d\lambda. \end{aligned}$$

Thus λ is the unique solution to the claim of the Lemma. This uniqueness together with $c^n_k = c^n_{n-k}$, implies that λ is symmetric. Thus any intricacy defines a measure as claimed.

We turn to the converse, considering a symmetric measure λ on $[0,1]$ and defining c by means of (3.5). The coefficients depending only on the cardinalities, $\mathcal{I}^c$ is trivially exchangeable. The symmetry of λ yields immediately $c^n_k = c^n_{n-k}$, and the normalization condition is given by

$$\sum_{k=0}^{N} \binom{N}{k} c^N_k = \int_{[0,1]} \sum_{k=0}^{N} \binom{N}{k} x^k (1-x)^{N-k}\, \lambda(dx) = 1,$$

i.e. $c \in \mathcal{C}(\mathbb{N}^*)$. To prove the projectivity of c, namely (3.7), we compute:

$$\begin{aligned}\sum_{\ell=0}^{M} c^{M+N}_{k+\ell} \binom{M}{\ell} &= \int_{[0,1]} \left[\sum_{\ell=0}^{M} \binom{M}{\ell} x^\ell (1-x)^{M-\ell} \right] x^k (1-x)^{N-k} \lambda(dx) \\ &= \int_{[0,1]} x^k (1-x)^{N-k} \lambda(dx) = c^N_k .\end{aligned}$$

Thus (3.7) and projectivity follow. The Lemma is proved. □

Proof of Proposition 3.3. First, let $\mathcal{I}^c$ be an intricacy. Lemma 3.3 yields a symmetric probability measure λ_c on $[0,1]$ satisfying (3.5). If W_c be a random variable with law λ_c, then (3.2) is equivalent to (3.5).

Conversely, suppose that $c = (c^I_S)_{S \subset I}$ has the form (3.2) for some probability $\mathbb{P}$ defined by $W_c, Y_1, Y_2, \dots$ as in the statement. Obviously $c^I_S \geq 0$ and $\sum_{S \subset I} c^I_S = 1$. $c^I_S = c^I_{S^c}$ follows from the symmetry of W_c. Thus c is a system of coefficients. Exchangeability of c follows from exchangeability of the random variables $\mathbb{1}_{(Y_i \geq W_c)}$, $i \in I$. By (3.2) we know that

$$c^I_S = c^{|I|}_{|S|} = \mathbb{E}\left((1-W_c)^{|I \setminus S|}\, W_c^{|S|} \right) = \int_{[0,1]} x^{|S|} (1-x)^{|I \setminus S|}\, \lambda_c(dx).$$

Therefore, by Lemma 3.5 the functional $\mathcal{I}^c$ is an intricacy.

Let now $W_c, Y_1, Y_2, \dots$ be defined as in point 2 of the statement and $\mathcal{Z}$ defined by (3.4). Each $i \in \mathbb{N}^*$ belongs to the random set $\mathcal{Z}$ if and only if $Y_i \geq W_c$. Conditionally on W_c, the probability of $\{Y_i \geq W_c\}$ is therefore $1 - W_c$. As the variables $Y_1, Y_2, \dots$ are independent:

$$\mathbb{P}(\mathcal{Z} \cap I = S \,|\, W_c) = (1-W_c)^{|I \setminus S|}\, W_c^{|S|}.$$

Averaging over the values of W_c we obtain

$$\mathbb{P}(\mathcal{Z} \cap I = S) = \mathbb{E}\left((1-W_c)^{|I \setminus S|}\, W_c^{|S|} \right)$$

and therefore (3.2) and (3.3) are equivalent. The last assertion follows from Lemma 3.5 and from (3.3). The Proposition is proved. □

3.3. **Examples of intricacies.** We show that the Edelman-Sporns-Tononi neural complexity (1.1) and the uniform and p-symmetric intricacies correspond to natural probability laws on $[0,1]$. In particular, they are weakly additive and really intricacies:

Lemma 3.7. *In the setting of Lemma 3.5*

(1) If W_c is uniform on $[0,1]$ then $\mathcal{I}^c$ is the Edelman-Sporns-Tononi neural complexity (1.1).
(2) If W_c is uniform on $\{p, 1-p\}$ then $\mathcal{I}^c$ is the p-symmetric intricacy $\mathcal{I}^p$; in the case $p = 1/2$, $W_c = \frac{1}{2}$ a.s. yields the uniform intricacy $\mathcal{I}^U$.

Proof. Let W_c be uniform on $[0,1]$. Then

$$\begin{aligned}\mathbb{P}(\mathcal{Z} \cap I = \{1, \ldots, k\}) &= \mathbb{P}(Z_1 = \cdots = Z_k = 1,\, Z_{k+1} = \cdots = Z_N = 0) \\ &= \int_{[0,1]} x^k (1-x)^{N-k} \, dx =: a(k, N-k).\end{aligned}$$

We claim now that for all $k \geq 1$ and $j \geq 0$

$$a(k,j) = \frac{j!}{(k+1)\cdots(k+j+1)} = \frac{1}{(k+j+1)\binom{k+j}{k}},$$

i.e., the Edelman-Sporns-Tononi coefficient c_j^{k+j}.

Indeed, for $j = 0$ this reduces to $\int_0^1 x^k \, dx = 1/(k+1)$. To prove the general case, one fixes k and uses recurrence on j. Indeed, suppose we have the result for $j \geq 0$. Then

$$\begin{aligned}&\int_0^1 x^k (1-x)^{j+1} \, dx = \int_0^1 x^k (1-x)^j \, dx - \int_0^1 x^{k+1} (1-x)^j \, dx \\ &= \frac{1}{(k+j+1)\binom{k+j}{k}} - \frac{1}{(k+j+2)\binom{k+j+1}{k+1}} = \frac{1}{(k+j+2)\binom{k+j+1}{k}}.\end{aligned}$$

If W_c is uniform over $\{p, 1-p\}$ then

$$\int_{[0,1]} x^k (1-x)^{N-k} \, \frac{1}{2}(\delta_p + \delta_{1-p})(dx) = \frac{1}{2}(p^k(1-p)^{N-k} + (1-p)^k p^{N-k}),$$

which is the coefficient c_k^N of $\mathcal{I}^p$. □

3.4. **Further properties.** We deduce some useful facts from the above representation.

Lemma 3.8. *The following are equivalent for an intricacy $\mathcal{I}^c$ with associated measure λ_c as in Lemma 3.5.*

(1) $\mathcal{I}^c$ is non-null, i.e. $c_k^N > 0$ for at least one choice of $N \geq 2$ and $1 \leq k < N$;
(2) $c_k^N > 0$ for all $N \geq 2$ and $1 \leq k \leq N-1$;
(3) $\lambda_c(]0,1[) > 0$.

Proof. We have:

$$c_j^n = \int_{[0,1]} x^j (1-x)^{n-j} \, \lambda_c(dx)$$

with $x^j(1-x)^{n-j}$ zero exactly at $x \in \{0,1\}$ whenever $0 < j < n$ and strictly positive on $]0,1[$. Thus (1) $\Longrightarrow$ (3) $\Longrightarrow$ (2) $\Longrightarrow$ (1). $\square$

Lemma 3.9. *If $\mathcal{I}^c$ is non-null, then $\mathcal{I}^c(X) = 0$ for a $X \in \mathcal{X}(d,N)$ if and only if $X = (X_1, \ldots, X_N)$ is an independent family.*

Proof. It is enough to show that: $\mathcal{I}^c(X) = 0 \iff \mathrm{H}(X) = \sum_{i \in I} \mathrm{H}(X_i)$. If $\mathcal{I}^c$ is non-null and $\mathcal{I}^c(X) = 0$, then by Lemma 3.8 we have $\mathrm{MI}(S) = 0$ for all $S \subset I$ with $S \notin \{\emptyset, I\}$. Therefore $\mathrm{H}(X) = \mathrm{H}(X_S) + \mathrm{H}(X_{S^c})$ and an easy induction yields the claim. $\square$

Example 3.10 (Systems with intricacy proportional to N)**.** We generalize the result of Example 2.14 from $\mathcal{I}^U$ to a non-null intricacy $\mathcal{I}^c$. Considering the same system X as in Example 2.14, we get by Proposition 3.3

$$\begin{aligned}\frac{\mathcal{I}^c(X)}{\log d} &= \sum_{S \subset I} c_S^I \, |\Delta_S| = \mathbb{E}\left(|\Delta_{\mathcal{Z} \cap I}|\right) \\ &= \sum_{k=1}^{N-1} \mathbb{P}(\mathbb{1}_{\mathcal{Z}}(k) \neq \mathbb{1}_{\mathcal{Z}}(k+1)) = (N-1)\, \mathbb{P}(\mathbb{1}_{\mathcal{Z}}(1) \neq \mathbb{1}_{\mathcal{Z}}(2)).\end{aligned}$$

By the probabilistic representation (3.2) through a random variable W_c with law λ_c on $[0,1]$,

$$\kappa_c := \mathbb{P}(\mathbb{1}_{\mathcal{Z}}(1) \neq \mathbb{1}_{\mathcal{Z}}(2)) = \int_{[0,1]} 2x(1-x)\, \lambda_c(dx) \in]0, 1/2]. \tag{3.9}$$

Then we have obtained a system $X \in \mathcal{X}(d^2, N)$ such that

$$\mathcal{I}^c(X) = \frac{\kappa_c}{2}\, (N-1) \log d^2. \qquad \square \tag{3.10}$$

4. Bounds for maximal intricacies

In this section we prove Theorem 1.3. We recall the definition (3.9) for a non-null intricacy $\mathcal{I}^c$

$$\kappa_c = 2 \int_{[0,1]} x(1-x)\, \lambda_c(dx) = 2c_1^2 > 0. \tag{4.1}$$

Recall that $\mathcal{I}^c(d,N)$ and $\mathcal{I}^c(d,N,x)$, defined in (1.4) and (1.5), denote the maximum of $\mathcal{I}^c$ over $\mathcal{M}(d,N)$, respectively over $\{\mu \in \mathcal{M}(d,N) : \mathrm{H}(\mu) = xN \log d\}$. We are going to show the following

Proposition 4.1. *Let $\mathcal{I}^c$ be a non-null intricacy and $d \geq 2$. Then for all $N \geq 2$*

$$\frac{\kappa_c \log d}{2} \left(1 - \frac{1}{N}\right) \leq \frac{\mathcal{I}^c(d,N)}{N} \leq \frac{\log d}{2}, \tag{4.2}$$

and for any $x \in [0,1]$

$$[x \wedge (1-x)]\ \kappa_c \log d \left(1 - \frac{1}{N}\right) \le \frac{\mathcal{I}^c(d,N,x)}{N} \le \frac{1}{2}\log d, \tag{4.3}$$

where $\kappa_c > 0$ *is defined in* (4.1).

Proof. The upper bound for $\mathcal{I}^c(d,N)/N$ follows from Lemma 2.6. We show now the lower bound for $\mathcal{I}^c(d,N,x)/N$. Let $x \in]0,1[$. In example 2.13 we have constructed a system $X = (X_1, X_2) \in \mathcal{X}(d,2)$ with

$$\mathrm{H}(X) = x \log d^2, \qquad \mathcal{I}^c(X) = 2\kappa_c\,[x \wedge (1-x)] \log d > 0.$$

Let now $(Y_{2i+1})_{i\ge 0}$ an i.i.d. family of copies of X_1 and set $Y_{2(i+1)} := Y_{2i+1}$ for all $i \ge 0$. Then, for $M \ge 1$, $Y := (Y_i)_{i=1,\dots,2M} \in \mathcal{X}(d,2M)$ is the product of M independent copies of (X_1, X_2) and by weak additivity

$$\mathcal{I}^c(Y) = M\,\mathcal{I}^c(X) = 2M\,\kappa_c\,[x \wedge (1-x)] \log d, \qquad \mathrm{H}(Y) = 2Mx \log d.$$

If S is a $\{0,\dots,d-1\}$-valued random variable independent of Y with $\mathrm{H}(Z) = x \log d$, then $Z := (Y_1, \dots, Y_{2M}, S) \in \mathcal{X}(d, 2M+1)$ satisfies by weak additivity

$$\mathcal{I}^c(Z) = \mathcal{I}^c(Y_1, \dots, Y_{2M}) = 2M\,\kappa_c\,[x \wedge (1-x)] \log d, \qquad \mathrm{H}(Z) = (2M+1)x \log d.$$

Setting $N = 2M$, respectively $N = 2M+1$, we obtain the upper bound for $\mathcal{I}^c(d,N,x)/N$. Taking the supremum over $x \in [0,1]$ in (4.3), we obtain (4.2). □

4.1. **Super-additivity.** We are going to prove that the maps $N \mapsto \mathcal{I}^c(d,N)$ and $N \mapsto \mathcal{I}^c(d,N,x)$ are super-additive. By Lemma 2.4, the suprema defining $\mathcal{I}^c(d,N)$ and $\mathcal{I}^c(d,N,x)$ are maxima. The measures achieving the first supremum are called *maximal intricacy measures.*

Lemma 4.2. *For any intricacy* $\mathcal{I}^c$ *and* $d \ge 2$*, the following limits exist. First,*

$$\mathcal{I}^c(d) = \lim_{N\to\infty} \frac{\mathcal{I}^c(d,N)}{N} = \sup_{N\ge 1} \frac{\mathcal{I}^c(d,N)}{N} \in\ ]0,+\infty[\tag{4.4}$$

and, for each $x \in]0,1[$,

$$\mathcal{I}^c(d,x) = \lim_{N\to\infty} \frac{\mathcal{I}^c(d,N,x)}{N} = \sup_{N\ge 1} \frac{\mathcal{I}^c(d,N,x)}{N} \in\ ]0,+\infty[. \tag{4.5}$$

Proof. We prove (4.5), (4.4) being similar and simpler. Fix $x \in]0,1[$. For each $N \ge 1$, let $a_N := \mathcal{I}^c(d,N,x)$. We claim that this sequence is *super-additive*, i.e.,

$$a_{N+M} \ge a_N + a_M, \qquad \forall\ N, M \ge 1.$$

Indeed, let X^N and X^M such that

$$\mathcal{I}^c(X^N) = \mathcal{I}^c(d,N,x), \quad \mathrm{H}(X^N) = xN \log d,$$
$$\mathcal{I}^c(X^M) = \mathcal{I}^c(d,M,x), \quad \mathrm{H}(X^M) = xM \log d.$$

Assume that X^N and X^M are independent. By weak-additivity

$$\mathcal{I}^c(X^N, X^M) = \mathcal{I}^c(X^N) + \mathcal{I}^c(X^M),$$
$$\mathrm{H}(X^N, X^M) = \mathrm{H}(X^N) + \mathrm{H}(X^M) = x(N+M) \log d.$$

Thus,

$$\begin{aligned} a_N + a_M &= \mathcal{I}^c(d, N, x) + \mathcal{I}^c(d, M, x) = \mathcal{I}^c\left(X^N\right) + \mathcal{I}^c\left(X^M\right) \\ &= \mathcal{I}^c\left(X^N, X^M\right) \le \mathcal{I}^c(d, N+M, x) = a_{N+M}. \end{aligned}$$

Moreover, by Proposition 4.1, we have $\sup_{N\ge 1} a_N/N \le (\log d)/2$. Therefore, by Fekete's Lemma $a_N/N \to \sup_M a_M/M \le (\log d)/2$ as $N \to +\infty$. Moreover, the limit is positive by (4.3). □

4.2. **Adjusting Entropy.** To strengthen the previous result to obtain the second assertion of Theorem 1.3, we must adjust the entropy without significantly changing the intricacy.

Lemma 4.3. *Let $X^{(1)}, \ldots, X^{(r)} \in \mathcal{X}(d, N)$. Let U be a random variable over $\{1, \ldots, r\}$, independent of $\{X^{(1)}, \ldots, X^{(r)}\}$. Let $Y := X^{(U)} \in \mathcal{X}(d, N)$, i.e., $Y = X^{(u)}$ whenever $U = u$. Then:*

$$0 \le \mathrm{H}(Y_S) - \sum_{u=1}^{r} \mathbb{P}(U = u)\, \mathrm{H}(X_S^{(u)}) \le \log r, \qquad \forall\, S \subset \{1, \ldots, N\}, \tag{4.6}$$

$$-\log r \le \mathcal{I}^c(Y) - \sum_{u=1}^{r} \mathbb{P}(U = u)\, \mathcal{I}^c(X^{(u)}) \le 2\log r. \tag{4.7}$$

Proof. We first prove (4.6). By (A.2),

$$\mathrm{H}(Y_S|\, U) \le \mathrm{H}(Y_S) \le \mathrm{H}(Y_S, U) = \mathrm{H}(Y_S|\, U) + \mathrm{H}(U).$$

Now $\mathrm{H}(U) \le \log r$. (4.6) follows as:

$$\mathrm{H}(Y_S|\, U) = \sum_{u=1}^{r} \mathbb{P}(U = u)\, \mathrm{H}(Y_S|\, U = u) = \sum_{u=1}^{r} \mathbb{P}(U = u)\, \mathrm{H}(X_S^{(u)}).$$

(4.7) follows immediately, using (2.2) and (4.6). □

Lemma 4.4. *Let $0 < x < 1$ and $\epsilon > 0$ and $\mathcal{I}^c$ be some non-null intricacy. Then there exists $\delta_0 > 0$ and $N_0 < \infty$ with the following property for all $0 < \delta < \delta_0$ and $N \ge N_0$. For any $X \in \mathcal{X}(d, N)$ such that $\left|\frac{\mathrm{H}(X)}{N\log d} - x\right| \le \delta$, there exists $Y \in \mathcal{X}(d, N)$ satisfying:*

$$\mathrm{H}(Y) = xN\log d, \qquad |\mathcal{I}^c(Y) - \mathcal{I}^c(X)| \le \epsilon N \log d.$$

Proof. We fix $\delta_0 = \delta_0(\epsilon, x) > 0$ so small that:

$$\frac{\delta_0}{\min\{1 - x - \delta_0, x - \delta_0\}} < \epsilon/4$$

and $N_0 = N_0(\epsilon, x, \delta_0)$ so large that:

$$\frac{\log 2}{N_0 \min\{1 - x - \delta_0, x - \delta_0\} \log d} < \epsilon/4.$$

Let $N \ge N_0$ and $X \in \mathcal{X}(d, N)$ be such that $\left|\frac{\mathrm{H}(X)}{N\log d} - x\right| \le \delta \le \delta_0$. There are two similar cases, depending on whether $\mathrm{H}(X)$ is greater or less than $xN\log d$.

We assume $h := \mathrm{H}(X)/N\log d < x$ and shall explain at the end the necessary modifications for the other case.

Let $Z = (Z_i, i = 1, \dots, N)$ be i.i.d. random variables, uniform over $\{0, \dots, d-1\}$. We consider $Y^t \in \mathcal{X}(d,N)$ defined by

$$Y^t := X\,\mathbb{1}_{(U\le 1-t)} + Z\,\mathbb{1}_{(U>1-t)},$$

where U is a uniform random variable over $[0,1]$ independent of X and Z. $\mathcal{I}^c(Y^0) = \mathcal{I}^c(X)$ and $\mathcal{I}^c(Y^1) = \mathcal{I}^c(Z) = 0$. Hence, by the continuity of the intricacy, we get that there is some $0 < t_0 < 1$ such that $\mathrm{H}(Y^{t_0}) = xN\log d$. Let us check that t_0 is small.

By (4.6)

$$0 \le \mathrm{H}(Y^t) - (1-t)\,\mathrm{H}(X) - t\,\mathrm{H}(Z) = \mathrm{H}(Y^t) - (1-t)hN\log d - tN\log d \le \log 2.$$

so that, for some $\alpha \in [0,1]$,

$$0 < t_0 = \frac{x-h}{1-h} - \frac{\alpha\log 2}{N(1-h)\log d} \le \frac{\delta}{1-x-\delta} < \frac{\epsilon}{2},$$

since $\delta \le \delta_0$. Thus, by (4.7), setting $Y := Y^{t_0}$,

$$|\mathcal{I}^c(Y) - (1-t_0)\mathcal{I}^c(X) - t_0\,\mathcal{I}^c(Z)| = |\mathcal{I}^c(Y) - (1-t_0)\mathcal{I}^c(X)| \le 2\log 2,$$

and therefore by (4.2)

$$|\mathcal{I}^c(Y) - \mathcal{I}^c(X)| \le t_0\mathcal{I}^c(X) + 2\log 2 \le \frac{\epsilon}{2}\,N\log d + 2\log 2.$$

Dividing by $N\log d \ge N_0\log d$ we obtain the desired estimate.

For the case $h > x$, we use instead a system Z with constant variables, so that $\mathrm{H}(Z) = 0 = \mathcal{I}^c(Z)$ and a similar argument gives the result. □

4.3. **Proof of Theorem 1.3.** Assertion (1) is already established: see Proposition 4.1. It remains to complete the proof of the second assertion.

Let us set for $\delta \ge 0$

$$\mathcal{I}^c(d,N,x,\delta) := \sup\left\{\mathcal{I}^c(X) : X \in \mathcal{X}(d,N),\ \left|\frac{\mathrm{H}(X)}{N\log d} - x\right| \le \delta\right\}.$$

We want to prove that

$$\mathcal{I}^c(d,x) = \lim_{N\to+\infty} \frac{1}{N}\,\mathcal{I}^c(d,N,x,\delta_N).$$

for any sequence $\delta_N \ge 0$ converging to 0 as $N \to +\infty$. We first observe that (4.5) gives that the limit exists and is equal to $\mathcal{I}^c(d,x)$ if $\delta_N = 0$, for all $N \ge 1$. Consider now a general sequence of non-negative numbers δ_N converging to zero. Obviously, $\mathcal{I}^c(d,N,x,\delta_N) \ge \mathcal{I}^c(d,N,x,0)$, so that

$$\liminf_{N\to\infty} \frac{1}{N}\left(\mathcal{I}^c(d,N,x,\delta_N) - \mathcal{I}^c(d,N,x,0)\right) \ge 0.$$

Let us prove the reverse inequality for the lim sup. Let $\epsilon > 0$. Let $X^N \in \mathcal{X}(d,N)$ realize $\mathcal{I}^c(d,N,x,\delta_N)$. Let δ_0 and N_0 be as in Lemma 4.4. We may assume that $N \ge N_0$ and $\delta_N < \delta_0$. It follows that there is some $Y^N \in \mathcal{X}(d,N)$ with the entropy

$Nx\log d$ such that $\mathcal{I}^c(Y^N) \geq \mathcal{I}^c(X^N) - \epsilon N$. Hence, $\mathcal{I}^c(d,N,x,0) \geq \mathcal{I}^c(d,N,x,\delta_N) - \epsilon N$. We obtain

$$\limsup_{N\to\infty} \frac{1}{N}\left(\mathcal{I}^c(d,N,x,\delta_n) - \mathcal{I}^c(d,N,x,0)\right) \leq \epsilon,$$

Assertion (2) follows by letting $\epsilon \to 0$. $\square$

5. Exchangeable systems

In this section we prove Theorem 1.5, namely we prove that exchangeable systems have small intricacy. In particular, one cannot approach the maximal intricacy with such systems.

Proposition 5.1. *Let $\mathcal{I}^c$ be any mutual information functional and $d \geq 2$. Then for all $\varepsilon > 0$ there exists a constant $C = C(\varepsilon, d)$ such that for all exchangeable $X \in \mathcal{X}(d,N)$*

$$\mathcal{I}^c(X) \leq CN^{\frac{2}{3}+\varepsilon}, \qquad N \geq 2. \tag{5.1}$$

In particular

$$\lim_{N\to\infty} \frac{1}{N} \max_{X\in \mathrm{EX}(d,N)} \mathcal{I}^c(X) = 0.$$

Proof. Fix $\varepsilon > 0$. Throughout the proof, we denote by C constants which only depend on d and ε and which may change value from line to line. Also $\mathbf{k} = (k_1, \dots, k_d) \in \mathbb{N}^d$, $\mathbf{x} := \frac{1}{n}\mathbf{k}$ and $|\mathbf{k}| := k_1 + \cdots + k_d = n$ and the multinomial coefficients and the entropy function are denoted by:

$$\binom{n}{\mathbf{k}} = \frac{n!}{k_1!k_2!\dots k_d!}, \qquad h(\mathbf{x}) = -\sum_{i=1}^{d} x_i \log x_i.$$

We are going to use the following version of Stirling's formula

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\zeta_n}, \qquad \frac{1}{12n+1} < \zeta_n < \frac{1}{12n}, \qquad n \geq 1.$$

Therefore, for all $\mathbf{k} \in \mathbb{N}^d$ such that $|\mathbf{k}| = n$

$$\binom{n}{\mathbf{k}} = \left[e^{nh(\mathbf{x})}(2\pi n)^{1/2}\prod_{x_i\neq 0}(2\pi n x_i)^{-1/2}\right] g(\mathbf{k}, n),$$

where $g(\mathbf{k}, n) := \exp(\zeta_n - \zeta_{k_1} - \cdots - \zeta_{k_d})$ and therefore

$$\exp(-d) \leq g(\mathbf{k}, n) \leq \exp(1).$$

In particular, as all non-zero x_i satisfy $x_i \geq 1/n$,

$$\left|\frac{1}{n}\log\binom{n}{\mathbf{k}} - h(\mathbf{x})\right| \leq C\,\frac{\log n}{n}. \tag{5.2}$$

Let $X \in \mathrm{EX}(d,N)$. We set for $0 \leq n \leq N$ and $|\mathbf{k}| = n$

$$p_{n,\mathbf{k}} = \mathbb{P}(X_1 = \cdots = X_{k_1} = 1, \dots, X_{k_1+\cdots+k_{d-1}+1} = \cdots = X_n = d).$$

These $\binom{n+d-1}{d-1}$ numbers determine the law of any subsystem X_S of size $|S|=n$. It is convenient to define also $Y_i := \#\{1 \le j \le n : X_j = i\}$ for $i=0,\dots,d-1$ and

$$q_{n,\mathbf{k}} := \mathbb{P}(Y_i = k_i,\ i=0,\dots,d-1) = \binom{n}{\mathbf{k}} p_{n,\mathbf{k}}.$$

Since the vector $(q_{n,\mathbf{k}})_{|\mathbf{k}|=n}$ gives the law of the vector $(Y_1,\dots,Y_d)$ we have in particular

$$\sum_{|\mathbf{k}|=n} q_{n,\mathbf{k}} = 1.$$

Second, we observe that for $|S|=n$

$$\left| \frac{\mathrm{H}(X_S)}{n} - \frac{1}{n} \sum_{|\mathbf{k}|=n} q_{n,\mathbf{k}}\, h(\mathbf{x}) \right| \le C\, \frac{\log n}{n}. \tag{5.3}$$

Indeed

$$\begin{aligned} \frac{\mathrm{H}(X_S)}{n} &= -\frac{1}{n} \sum_{|\mathbf{k}|=n} q_{n,\mathbf{k}} \,\log \frac{q_{n,\mathbf{k}}}{\binom{n}{\mathbf{k}}} = \sum_{|\mathbf{k}|=n} q_{n,\mathbf{k}}\, \frac{1}{n}\, \log \binom{n}{\mathbf{k}} - \frac{1}{n} \sum_{|\mathbf{k}|=n} q_{n,\mathbf{k}} \log q_{n,\mathbf{k}} \\ &= \frac{1}{n} \sum_{|\mathbf{k}|=n} q_{n,\mathbf{k}}\, h(\mathbf{x}) + G(n), \qquad |G(n)| \le C\, \frac{\log n}{n}, \end{aligned}$$

where we use (5.2) and the fact that

$$- \sum_{|\mathbf{k}|=n} q_{n,\mathbf{k}} \log q_{n,\mathbf{k}} = \mathrm{H}(Y_1,\dots,Y_d) \le d \log n,$$

since the support of the random vector $(Y_1,\dots,Y_d)$ has cardinality at most n^d.

Third, we claim that, for $\varepsilon > 0$, there exists a constant C such that for all N and all $X \in \mathrm{EX}(d,N)$, for all $n \in [\tilde N, N]$ with $\tilde N := \lfloor N^{\frac{2}{3}+\varepsilon} + 1 \rfloor$,

$$\left| \sum_{|\mathbf{k}|=n} q_{n,\mathbf{k}} h(\mathbf{x}) - \sum_{|\mathbf{K}|=N} q_{N,\mathbf{K}} h(\mathbf{X}) \right| \le C\, N^{-\frac{1}{3}+\varepsilon}, \tag{5.4}$$

where $\mathbf{X} := \frac{1}{N}\mathbf{K}$ (no relation with the random variable X). By (5.3) and (5.4) we obtain for all $n \in [\tilde N, N]$ and $|S|=n$

$$\left| \frac{\mathrm{H}(X_S)}{n} - \frac{\mathrm{H}(X)}{N} \right| \le C\, N^{-\frac{1}{3}+\varepsilon}. \tag{5.5}$$

Let us show how (5.5) implies (5.1). Using $\mathrm{H}(X_S) \le \log d \cdot |S|$, $\sum_{S\subset I} c_S^I = 1$, we get

$$\sum_{|S|<\tilde N} c_S^I \,\mathrm{MI}(S) \le \sum_{S\subset I} c_S^I \times \log d \cdot \tilde N = \log d \cdot \tilde N.$$

Using (2.2), exchangeability of X, $\sum_{n=0}^{N} c_n^N \binom{N}{n} = 1$ and (5.5), we estimate

$$\begin{aligned}\mathcal{I}^c(X) &\le 2 \cdot \log d \cdot \tilde{N} + 2 \sum_{n=\tilde{N}}^{N} \binom{N}{n} c_n^N \, \mathrm{H}(X_{\{1,\dots,n\}}) - \mathrm{H}(X) \\ &\le 2 \sum_{n=0}^{N} \binom{N}{n} c_n^N \, n \left(\frac{\mathrm{H}(X)}{N} + C\, N^{-\frac{1}{3}+\varepsilon} \right) - \mathrm{H}(X) + C\tilde{N}.\end{aligned}$$

Finally, using $c_n^N \binom{N}{n} = c_{N-n}^N \binom{N}{N-n}$ and $\sum_{n=0}^{N} c_n^N \binom{N}{n} = 1$

$$\begin{aligned}\mathcal{I}^c(X) &\le \left(2 \sum_{n=0}^{N} c_n^N \binom{N}{n} \frac{n}{N} - 1 \right) \mathrm{H}(X) + CN \times N^{-\frac{1}{3}+\varepsilon} + C\tilde{N} \\ &\le \left(\sum_{n=0}^{N} c_n^N \binom{N}{n} \left(\frac{n}{N} + \frac{N-n}{N} \right) - 1 \right) \mathrm{H}(X) + CN^{\frac{2}{3}+\varepsilon} = CN^{\frac{2}{3}+\varepsilon}\end{aligned}$$

and (5.1) is proved.

We turn now to the proof of (5.4). We claim first that

$$p_{n,\mathbf{k}} = \sum_{|\mathbf{K}|=N,\, \mathbf{K}\ge\mathbf{k}} p_{N,\mathbf{K}} \binom{N-n}{\mathbf{K}-\mathbf{k}}. \tag{5.6}$$

Indeed, notice that

$$p_{n,\mathbf{k}} = \sum_{j=1}^{d} p_{n+1,\mathbf{k}+\delta^j}, \qquad \forall\, 0 \le n < N,\ \forall\, |\mathbf{k}| = n,$$

where $\delta^j := (\delta_1^j, \dots, \delta_d^j)$ with $\delta_i^j = 1$ if $i = j$, 0 otherwise. This in particular yields (5.6) for $N = n+1$. Moreover if $|\mathbf{K}| = n+1$ then

$$\binom{n+1}{\mathbf{K}} = \sum_{j=1}^{d} \binom{n}{\mathbf{K}-\delta^j} \mathbb{1}_{(\mathbf{K}\ge\delta^j)}.$$

Then, arguing by induction on $N \ge n$

$$\begin{aligned}p_{n,\mathbf{k}} &= \sum_{|\mathbf{K}|=N,\, \mathbf{K}\ge\mathbf{k}} p_{N,\mathbf{K}} \binom{N-n}{\mathbf{K}-\mathbf{k}} = \sum_{|\mathbf{K}|=N,\, \mathbf{K}\ge\mathbf{k}} \sum_{j=1}^{d} p_{N+1,\mathbf{K}+\delta^j} \binom{N-n}{\mathbf{K}-\mathbf{k}} \\ &= \sum_{|\mathbf{K}'|=N+1} p_{N+1,\mathbf{K}'} \sum_{j=1}^{d} \binom{N-n}{\mathbf{K}'-\mathbf{k}-\delta^j} \mathbb{1}_{(\mathbf{K}-\mathbf{k}\ge\delta^j)} \\ &= \sum_{|\mathbf{K}'|=N+1,\mathbf{K}'\ge\mathbf{k}} p_{N+1,\mathbf{K}'} \binom{N+1-n}{\mathbf{K}'-\mathbf{k}}.\end{aligned}$$

We recall that $q_{n,\mathbf{k}} = \binom{n}{\mathbf{k}} p_{n,\mathbf{k}}$. Notice that it is enough to prove claim (5.4) in the case $q_{N,\mathbf{k}'} = \delta_{\mathbf{k}',\mathbf{K}}$, i.e., $p_{N,\mathbf{k}'} = \binom{N}{\mathbf{k}'}^{-1}$ for $\mathbf{k}' = \mathbf{K}$ and zero otherwise, if we find a constant C which does not depend on $(N, n, \mathbf{K})$. Indeed, the two expressions are linear and the average of $CN^{-1/3+\varepsilon}$ will remain of the same order. Thus, we need to estimate:

$$a(N,\mathbf{K},n,\mathbf{k}) := q_{n,\mathbf{k}} = \binom{n}{\mathbf{k}} \times \binom{N}{\mathbf{K}}^{-1} \binom{N-n}{\mathbf{K}-\mathbf{k}}.$$

Let $\mathbf{x} := \mathbf{k}/n \in [0,1]^d$, $\mathbf{X} := \mathbf{K}/N \in [0,1]^d$ and $\nu =: n/(N-n)$. Formula (5.2) implies that $\frac{1}{n}\log a(N,\mathbf{K},n,\mathbf{k})$ is equal to:

$$\underbrace{h(\mathbf{x}) - (1+\nu^{-1})h(\mathbf{X}) + \nu^{-1}h(\mathbf{X}+\nu(\mathbf{X}-\mathbf{x}))}_{=:\phi_{\nu,\mathbf{X}}(\mathbf{x})} + G(N,n),$$

where $|G(N,n)| \le \kappa(\log N)/n$, for some $\kappa = \kappa(d)$.

Let us now write for all $(x_1,\ldots,x_{d-1}) \in [0,1]^{d-1}$ such that $\sum_i x_i \le 1$

$$H(x_1,\ldots,x_{d-1}) := h(x_1,\ldots,x_d), \quad x_d := 1 - x_1 - \ldots - x_{d-1}.$$

Observe that for $i,j \le d-1$

$$\frac{\partial H}{\partial x_i} = \log\left(\frac{x_d}{x_i}\right), \qquad \frac{\partial^2 H}{\partial x_i \partial x_j} = -\frac{1}{x_d} - \frac{1}{x_i}\,\mathbb{1}_{(i=j)}.$$

In particular the Hessian of H is negative-definite, since for all $a \in \mathbb{R}^{d-1}\backslash\{0\}$

$$\sum_{i,j=1}^{d-1} a_i a_j \frac{\partial^2 H}{\partial x_i \partial x_j} = -\frac{1}{x_d}\left(\sum_{i=1}^{d-1} a_i\right)^2 - \sum_{i=1}^{d-1}\frac{1}{x_i}\,a_i^2 \le -\sum_{i=1}^{d-1} a_i^2$$

where we use the fact that $x_i \le 1$. Hence, h is concave and we obtain

$$\phi_{\nu,\mathbf{X}}(\mathbf{x}) = \frac{\nu+1}{\nu}\left[\frac{\nu}{\nu+1}\,h(\mathbf{x}) + \frac{1}{\nu+1}\,h((1+\nu)\mathbf{X} - \nu\mathbf{x}) - h(\mathbf{X})\right] \le 0,$$

so that the maximum of $\phi_{\nu,\mathbf{X}}$ is $0 = \phi_{\nu,\mathbf{X}}(\mathbf{X})$. The second order derivative estimate gives:

$$\phi_{\nu,\mathbf{X}}(\mathbf{x}) \le -2\|\mathbf{x}-\mathbf{X}\|^2 \qquad \text{where } \|\mathbf{x}\| := \sqrt{x_1^2 + \cdots + x_d^2}.$$

Combining with the bound $|G(N,n)| \le \kappa(\log N)/n$ above, we get, for all $n < N$:

$$a(N,\mathbf{K},n,\mathbf{k}) \le N^\kappa \times e^{-2n\|\mathbf{x}-\mathbf{X}\|^2}.$$

Recall $n \ge \tilde{N} = N^{\frac{2}{3}+\varepsilon}$ and set $\delta := N^{-\frac{1}{3}}$ and

$$\omega := \sup_{\|\mathbf{x}-\mathbf{X}\|<\delta} \|h(\mathbf{X}) - h(\mathbf{x})\| \le C\,\delta\log\frac{1}{\delta}.$$

Finally, using $h(\mathbf{x}) \le \log d$,

$$\left|\sum_{|\mathbf{k}|=n} q_{n,\mathbf{k}} h(\mathbf{x}) - h(\mathbf{X})\right| \le \omega \sum_{\|\mathbf{x}-\mathbf{X}\|<\delta} q_{n,\mathbf{k}} + 2\log d \sum_{\|\mathbf{x}-\mathbf{X}\|\ge\delta} q_{n,\mathbf{k}}$$

$$\le C\,\delta\log\frac{1}{\delta} + C\,n^d\,N^\kappa\,e^{-2\tilde{N}\delta^2} \le C(\log N)N^{-\frac{1}{3}} + CN^{\kappa+d}e^{-2N^\varepsilon} \le CN^{-\frac{1}{3}+\varepsilon}.$$

Then (5.4) and the proposition are proved. $\square$

6. Small support

In this section we prove Theorem 1.6, namely we show that exact maximizers have small support. Numerical experiments suggest that this support has in fact cardinality of order $d^{N/2}$. We are able to prove the following weaker estimate. For a fixed law $\mu \in \mathcal{M}(d,N)$, we call forbidden configurations the elements of $\Lambda_{d,N} := \{0,\dots,d-1\}^N$ with zero μ-probability.

Proposition 6.1. *Let $\mathcal{I}^c(X)$ be a non-null intricacy. Let $d=2$ and N large enough. Let $\mu \in \mathcal{X}(d,N)$ be a maximizer of $\mathcal{I}^c$. The forbidden configurations are a lower-bounded fraction of all configurations:*

$$\#\{\omega \in \Lambda_{d,N} : \mu([\omega]) = 0\} \geq c(d)|\Lambda_{d,N}|,$$

for some $c(d) > 0$ independent of N.

Proof. If $\mathcal{I}^c$ is non-null, then $\lambda_c(\{0,1\}) = 2\lambda_c(\{0\}) < 1$ and therefore $\lambda_c(\{0\}) < 1/2$. However we can without loss of generality suppose that $\lambda_c(\{0\}) = 0$: indeed it is enough to remark that

(1) the probability measure $\lambda_0 := \frac{\delta_0+\delta_1}{2}$ is associated with the null intricacy $\mathcal{I}^0 \equiv 0$,
(2) the correspondence $\lambda_c \mapsto \mathcal{I}^c$ is linear and one-to-one,
(3) we can write $\lambda_c = \alpha\lambda_0 + (1-\alpha)\lambda_{c'}$, where

$$\alpha := 2\lambda_c(\{0\}) < 1, \qquad \lambda_{c'}([a,b]) = \frac{\lambda_c([a,b]\cap]0,1[)}{\lambda_c(]0,1[)}, \quad \forall\, a \leq b.$$

Therefore $\mathcal{I}^c = \alpha\mathcal{I}^0 + (1-\alpha)\mathcal{I}^{c'} = (1-\alpha)\mathcal{I}^{c'}$ and $\mathcal{I}^{c'}$ has the same maximizers as $\mathcal{I}^c$ but with $\lambda_{c'}(\{0\}) = 0$

We fix some large integer z (how large will be explained below), $N > z$ and $d \geq 2$ and we consider the intricacy $\mathcal{I}^c$ as a function defined on the simplex $\mathcal{M}(d,N) = \{(p_\omega)_{\omega\in\Lambda_{d,N}} \in \mathbb{R}_+^{d^N} : \sum_{\omega\in\Lambda_{d,N}} p_\omega = 1\}$. A straightforward computation yields:

$$\frac{\partial \mathcal{I}^c}{\partial p_\omega} = -2\sum_{S\subset I} c_S^I \log\left(\sum_{\alpha\equiv\omega[S]} p_\alpha\right) + \log p_\omega - 1$$

where $\alpha \equiv \omega[S]$ iff $\alpha_i = \omega_i$ for all $i \in S$. The second derivatives are:

$$\frac{\partial^2 \mathcal{I}^c}{\partial p_\omega^2} = -2\sum_{S\subset I} \frac{c_S^I}{\sum_{\alpha\equiv\omega[S]} p_\alpha} + \frac{1}{p_\omega}, \qquad \frac{\partial^2 \mathcal{I}^c}{\partial p_{\omega_0}\partial p_{\omega_1}} = -2\sum_{S\subset I} \frac{c_S^I}{\sum_{\alpha\equiv\omega_0[S]} p_\alpha} \mathbb{1}_{(\omega_0=\omega_1[S])},$$

for $\omega_0 \neq \omega_1$.

Let $p = (p_\omega)_{\omega\in\Lambda_{d,N}}$ be a maximizer of $\mathcal{I}^c$. We show that for each $\beta \in \{0,\dots,d-1\}^{N-z}$,

$$\Omega_\beta := \left\{(\alpha_1,\dots,\alpha_z,\beta_1,\dots,\beta_{N-z}) \in \{0,\dots,d-1\}^N : \alpha \in \{0,\dots,d-1\}^z\right\}$$

must contain at least one configuration forbidden by p. The claim will follow since the cardinality of $\{0,\dots,d-1\}^{N-z}$ is d^N/d^z.

We assume by contradiction the existence of some $\beta \in \{0, \dots, d-1\}^{N-z}$ such that no configuration in Ω_β is forbidden. Let $\omega_0 \in \Omega_\beta$ be such that

$$p_{\omega_0} := \min\{p_\omega : \omega \in \Omega_\beta\} > 0.$$

Let now $\omega_1 \in \Omega_\beta \setminus \{\omega_0\}$, which exists since $|\Omega_\beta| \geq d \geq 2$, so that $p_{\omega_1} \geq p_{\omega_0} > 0$. We set for $t \in]-\varepsilon, \varepsilon[$ and $0 < \varepsilon < p_{\omega_0}$

$$p^t_\omega := \begin{cases} p_{\omega_1} + t, & \omega = \omega_1, \\ p_{\omega_0} - t, & \omega = \omega_0, \\ p_\omega, & \omega \notin \{\omega_0, \omega_1\}. \end{cases}$$

Then p^t is still a probability measure for $t \in]-\varepsilon, \varepsilon[$, since $p_{\omega_1} \geq p_{\omega_0} > \varepsilon > 0$.

Since p is a maximizer, then $\varphi(t) := \mathcal{I}^c(p^t) \leq \varphi(0) := \mathcal{I}^c(p)$ for $t \in]-\varepsilon, \varepsilon[$. Then

$$\begin{aligned} 0 \geq \varphi''(0) &= \frac{\partial^2 \mathcal{I}^c}{\partial p^2_{\omega_0}} + \frac{\partial^2 \mathcal{I}^c}{\partial p^2_{\omega_1}} - 2\frac{\partial^2 \mathcal{I}^c}{\partial p_{\omega_0} \partial p_{\omega_1}} \\ &= \frac{1}{p_{\omega_1}} + \frac{1}{p_{\omega_0}} - 2\sum_{S \subset I} \mathbb{1}_{(\omega_0 \neq \omega_1[S])} \left[\frac{c^I_S}{\sum_{\alpha \in [\omega_0]_S} p_\alpha} + \frac{c^I_S}{\sum_{\alpha \in [\omega_1]_S} p_\alpha} \right] \end{aligned}$$

where $[\omega]_S = \{\alpha : \alpha = \omega \mod [S]\}$ is the equivalence class of ω. Therefore

$$0 \geq \frac{1}{p_{\omega_1}} + \frac{1}{p_{\omega_0}} \left(1 - 2\sum_{S \subset I} \frac{c^I_S}{|[\omega_0]_S \cap \Omega_\beta|} - 2\sum_{S \subset I} \frac{c^I_S}{|[\omega_1]_S \cap \Omega_\beta|} \right)$$

and for some $\omega \in \Omega_\beta$

$$\sum_{S \subset I} \frac{c^I_S}{|[\omega]_S \cap \Omega_\beta|} > \frac{1}{4}. \tag{6.1}$$

On the other hand, we have:

$$|[\omega]_S \cap \Omega_\beta| = d^{|S^c \cap \{1,\dots,z\}|}$$

so that by Proposition 3.3 the left hand side of (6.1) is equal to:

$$\begin{aligned} \mathbb{E}\left(d^{-|\mathcal{Z}^c \cap \{1,\dots,z\}|}\right) &= \int_{[0,1]} \lambda_c(dx)\, \mathbb{E}\left(\prod_{i=1}^z d^{-\mathbb{1}_{(Y_i < x)}} \right) = \int_{[0,1]} \lambda_c(dx) \left(\frac{x}{d} + (1-x) \right)^z \\ &= \lambda_c(\{0\}) + \int_{]0,1]} \lambda_c(dx) \left(\frac{x}{d} + (1-x) \right)^z. \end{aligned}$$

Since we have reduced above to the case $\lambda_c(\{0\}) = 0$, then the latter expression tends to 0 as $z \to +\infty$, contradicting (6.1). $\square$

Appendix A. Entropy

In this Appendix, we recall needed facts from basic information theory. The main object is the entropy functional which may be said to quantify the randomness of a random variable.

Let X be a random variable taking values in a finite space E. We define the *entropy* of X

$$\mathrm{H}(X) := -\sum_{x\in E} P_X(x)\,\log(P_X(x)), \qquad P_X(x) := \mathbb{P}(X=x),$$

where we adopt the convention

$$0\cdot\log(0) = 0\cdot\log(+\infty) = 0.$$

We recall that

$$0 \le \mathrm{H}(X) \le \log|E|, \tag{A.1}$$

More precisely, $\mathrm{H}(X)$ is minimal iff X is a constant, it is maximal iff X is uniform over E. To prove (A.1), just notice that since $\varphi \ge 0$ and $\varphi(x) = 0$ if and only if $x \in \{0,1\}$, and by strict convexity of $x \mapsto \varphi(x) = x\log x$ and Jensen's inequality

$$\begin{aligned}\log|E| - \mathrm{H}(X) &= \frac{1}{|E|}\sum_{x\in E} P_X(x)\,|E|\,(\log(P_X(x)) + \log|E|)\\ &= \frac{1}{|E|}\sum_{x\in E}\varphi\left(P_X(x)\,|E|\right) \ge \varphi\left(\frac{1}{|E|}\sum_{x\in E} P_X(x)\,|E|\right) = \varphi(1) = 0,\end{aligned}$$

with $\log|E| - \mathrm{H}(X) = 0$ if and only if $P_X(x)\,|E|$ is constant in $x\in E$.

If we have a E-valued random variable X and a F-valued random variable Y defined on the same probability space, with E and F finite, we can consider the vector (X,Y) as a $E\times F$-valued random variable The entropy of (X,Y) is then

$$\mathrm{H}(X,Y) := -\sum_{x,y} P_{(X,Y)}(x,y)\,\log(P_{(X,Y)}(x,y)), \quad P_{(X,Y)}(x,y) := \mathbb{P}(X=x, Y=y).$$

This entropy $\mathrm{H}(X,Y)$ is a measure of the extent to which the "randomness of the two variables is shared". The following notions formalize this idea.

A.1. **Condidional Entropy.** The *conditional entropy* of X given Y is:

$$\mathrm{H}(X\,|\,Y) := \mathrm{H}(X,Y) - \mathrm{H}(Y).$$

We claim that

$$0 \le \mathrm{H}(X\,|\,Y) \le \mathrm{H}(X) \le \mathrm{H}(X,Y). \tag{A.2}$$

Remark that $P_X(x)$ and $P_Y(y)$, defined in the obvious way, are the marginals of $P_{(X,Y)}(x,y)$, i.e.

$$P_X(x) = \sum_y P_{(X,Y)}(x,y), \qquad P_Y(y) = \sum_x P_{(X,Y)}(x,y).$$

In particular, $P_X(x) \ge P_{(X,Y)}(x,y)$ for all x,y. Therefore

$$\sum_{x,y} P_{(X,Y)}(x,y)\,\log\left(\frac{P_{(X,Y)}(x,y)}{P_X(x)}\right) \le 0$$

which yields

$$\mathrm{H}(X,Y) = -\sum_{x,y} P_{(X,Y)}(x,y)\,\log P_{(X,Y)}(x,y) \geq -\sum_{x} P_X(x)\,\log P_X(x) = \mathrm{H}(X),$$

i.e. $\mathrm{H}(X,Y) \geq \mathrm{H}(X)$ and $\mathrm{H}(X|Y) \geq 0$. Therefore

$$\mathrm{H}(X,Y) \geq \max\{\mathrm{H}(X), \mathrm{H}(Y)\}. \tag{A.3}$$

Moreover $\mathrm{H}(X,Y) = \mathrm{H}(X)$, i.e. $\mathrm{H}(Y|X) = 0$, if and only if $P_{(X,Y)}(x,y) = P_X(x)$ whenever $P_{(X,Y)}(x,y) \neq 0$, i.e. Y is a function of X. On the other hand,

$$\mathrm{H}(X,Y) \leq \mathrm{H}(X) + \mathrm{H}(Y) \tag{A.4}$$

with equality, i.e., $\mathrm{H}(Y|X) = \mathrm{H}(Y)$, if and only if X and Y are independent. This shows that $\mathrm{H}(X\,|\,Y) \leq \mathrm{H}(X)$ and completes the proof of (A.2). Formula (A.4) can be shown by considering the Kullback-Leibler divergence or relative entropy:

$$I := \sum_{x,y} P_{(X,Y)}(x,y)\,\log\left(\frac{P_{(X,Y)}(x,y)}{P_X(x)\,P_Y(y)}\right).$$

Since $\log(\cdot)$ is concave, by Jensen's inequality

$$-I \leq \log\left(\sum_{x,y} P_{(X,Y)}(x,y)\,\frac{P_X(x)\,P_Y(y)}{P_{(X,Y)}(x,y)}\right) = \log\left(\sum_{x,y} P_X(x)\,P_Y(y)\right) = 0.$$

By strict concavity, $I = 0$ if and only if $P_{(X,Y)}(x,y) = P_X(x)\,P_Y(y)$ for all x,y, i.e., whenever X and Y are independent.

By the above considerations, $\mathrm{H}(X\,|\,Y) \in [0, \mathrm{H}(X)]$ is a measure of the uncertainty associated with X if Y is known. It is minimal iff X is a function of Y and it maximal iff X and Y are independent.

A.2. **Adding information decreases uncertainty.** Let us consider three random variables $(X,Y,Z) \mapsto E \times F \times G$ with E, F, G finite. Then we have that

$$\mathrm{H}(X\,|\,(Y,Z)) \leq \mathrm{H}(X\,|\,Y). \tag{A.5}$$

Indeed, this is equivalent to

$$\mathrm{H}(X,Y,Z) + \mathrm{H}(Y) \leq \mathrm{H}(X,Y) + \mathrm{H}(Y,Z).$$

Consider the quantity

$$J := \sum_{x,y,z} P_{(X,Y,Z)}(x,y,z)\,\log\left(\frac{P_{(X,Y,Z)}(x,y,z)\,P_Y(y)}{P_{(X,Y)}(x,y)\,P_{(Y,Z)}(y,z)}\right).$$

Since $-\log(\cdot)$ is convex, by Jensen's inequality

$$J \geq -\log\left(\sum_{x,y} \frac{P_{(X,Y)}(x,y)\,\sum_z P_{(Y,Z)}(y,z)}{P_Y(y)}\right) = -\log\left(\sum_{x,y} P_{(X,Y)}(x,y)\right) = 0,$$

and the inequality follows.

A.3. **Mutual Information.** Finally, we recall the notion of *mutual information* between two random variables X and Y defined on the same probability space:

$$\begin{aligned}\mathrm{MI}(X,Y) &:= \mathrm{H}(X)+\mathrm{H}(Y)-\mathrm{H}(X,Y)\\ &= \mathrm{H}(X)-\mathrm{H}(X\mid Y)=\mathrm{H}(Y)-\mathrm{H}(Y\mid X)\\ &= \sum_{x,y} P_{(X,Y)}(x,y)\,\log\left(\frac{P_{(X,Y)}(x,y)}{P_X(x)\,P_Y(y)}\right).\end{aligned}$$

This quantity is a measure of the common randomness of X and Y. By (A.3) and (A.4) we have $\mathrm{MI}(X,Y)\in[0,\min\{\mathrm{H}(X),\mathrm{H}(Y)\}]$. $\mathrm{MI}(X,Y)$ is minimal (zero) iff X,Y are independent and maximal, i.e. equal to $\min\{\mathrm{H}(X),\mathrm{H}(Y)\}$, iff one variable is a function of the other.

Mutual information is non-decreasing. Let $X,X',Y,Y',\hat{X},\hat{Y}$ be random variables such that X,X', resp. Y,Y', are (deterministic) functions of $\hat{X}$, resp. $\hat{Y}$. Then:

$$\mathrm{MI}(X,Y)\le \mathrm{MI}(\hat{X},\hat{Y}). \tag{A.6}$$

Mutual information is almost additive:

$$|\mathrm{MI}((X,Y),(X',Y'))-(\mathrm{MI}(X,X')+\mathrm{MI}(Y,Y'))|\le \mathrm{MI}(\hat{X},\hat{Y}). \tag{A.7}$$

These properties follow from the properties of conditional entropy. First,

$$\begin{aligned}&\mathrm{MI}(\hat{X},\hat{Y})=\mathrm{H}(\hat{X})+\mathrm{H}(\hat{Y})-\mathrm{H}(\hat{X},\hat{Y})\\ &\quad=\mathrm{H}(X)+\mathrm{H}(\hat{X}|X)+\mathrm{H}(Y)+\mathrm{H}(\hat{Y}|Y)-\mathrm{H}(X,Y)-\mathrm{H}(\hat{X}|X,Y)-\mathrm{H}(\hat{Y}|\hat{X},Y)\\ &\qquad=\mathrm{MI}(X,Y)+(\mathrm{H}(\hat{X}|X)-\mathrm{H}(\hat{X}|X,Y))+(\mathrm{H}(\hat{Y}|Y)-\mathrm{H}(\hat{Y}|\hat{X},Y)).\end{aligned}$$

(A.6) now follows from (A.5). Second,

$$\begin{aligned}\mathrm{MI}((X,Y),(X',Y'))&=\mathrm{H}(X,Y)+\mathrm{H}(X',Y')-\mathrm{H}(X,X',Y,Y')\\ &=\mathrm{H}(X)+\mathrm{H}(Y)-\mathrm{MI}(X,Y)+\mathrm{H}(X')+\mathrm{H}(Y')-\mathrm{MI}(X',Y')\\ &\quad-\mathrm{H}(X,X')-\mathrm{H}(Y,Y')+\mathrm{MI}((X,X'),(Y,Y'))\\ &=\mathrm{H}(X)+\mathrm{H}(X')-\mathrm{H}(X,X')+\mathrm{H}(Y)+\mathrm{H}(Y')-\mathrm{H}(Y,Y')\\ &\quad+(\mathrm{MI}((X,X'),(Y,Y'))-\mathrm{MI}(X,Y)-\mathrm{MI}(X',Y'))\\ &=\mathrm{MI}(X,X')+\mathrm{MI}(Y,Y')+(\mathrm{MI}((X,X'),(Y,Y'))-\mathrm{MI}(X,Y)-\mathrm{MI}(X',Y')).\end{aligned}$$

The nonnegativity of mutual information and (A.6) yields

$$\begin{aligned}-\min(\mathrm{MI}(X,Y),\mathrm{MI}(X',Y'))&\le \mathrm{MI}((X,Y),(X',Y'))-(\mathrm{MI}(X,X')+\mathrm{MI}(Y,Y'))\\ &\le \mathrm{MI}((X,X'),(Y,Y')).\end{aligned}$$

(A.7) follows.

References

[1] P. Bak, M. Paczuski, Complexity, contingency and criticality, *Proc. Natl. Acad. Sci. USA* **92** (1995), 6689-6696.

[2] L. Barnett, C. L. Buckley, S. Bullock, Neural complexity and structural connectivity, *Phys. Rev. E* **79** (2009), 051914.

[3] C. Bennett, How to define complexity in physics and why, in: *Complexity, entropy and the physics of information,* vol. VIII, eq. W. Zurek, Addison-Wesley (1990).

[4] J. Bertoin, *Random Fragmentation And Coagulation Processes*, Cambridge University Press (2006).

[5] J. Buzzi, L. Zambotti, *Approximate Maximizers of Intricacy Functionals,* preprint (2009), http://arxiv.org/abs/0909.2120.

[6] T. Cover, J. Thomas, *Elements of Information Theory,* John Wiley & Sons, 2006.

[7] J. Crutchfield, K. Young, Inferring statistical complexity, *Phys. Rev. Lett.* **63** (1989), 105–109.

[8] M. De Lucia, M. Bottaccio, M. Montuori, L. Pietronero, A topological approach to neural complexity, *Phys. Rev. E* **71** (2005), 016114 — arXiv:nlin/0411011v1.

[9] W. Feller, *An introduction to probability theory and its applications. Vol. 2,* John Wiley & Sons, 1971.

[10] G. Edelman, J. Gally, Degeneracy and complexity in biological systems, *Proc. Natl. Acad. Sci. USA,* **98** (2001), 13763–13768.

[11] N. Goldenfeld, L. Kadanoff, Simple lessons from complexity, Science **284** (1999), 87–89.

[12] A. Greven, G. Keller, G. Warnecke, *Entropy,* Princeton University Press, 2003.

[13] K. Holthausen, O. Breidbach, Analytical description of the evolution of neural networks: learning rules and complexity, Biol. Cybern. **81** (1999), 169–176.

[14] J. Krichmar, D. Nitz, J. Gally, G. Edelman, Characterizing functional hippocampal pathways in a brain-based device as it solves a spatial memory task, *Proc. Natl. Acad. Sci. USA,* **102** (2005), 2111–2116.

[15] A. Seth, E. Izhikevich, G. Reeke, G. Edelman, Theories and measures of consciousness: an extended framework, *Proc. Natl. Acad. Sci. USA,* **103** (2006), 10799–10804.

[16] Anil K Seth (2007), Models of consciousness, Scholarpedia, 2(1):1328.

[17] M. P. Shanahan, Dynamical complexity in small-world networks of spiking neurons, Phys. Rev. E 78, 041924 (2008).

[18] O. Sporns, G. Tononi, G. Edelman, Connectivity and complexity: the relationship between neuroanatomy and brain dynamics, Neural Netw. 2000 Oct-Nov;13(8-9):909-22.

[19] O. Sporns, Networks analysis, complexity, and brain function, *Complexity* **8** (2002), 56 - 60.

[20] O. Sporns (2007), Complexity, Scholarpedia, 2(10):1623.

[21] M. Talagrand, *Spin Glasses: A Challenge for Mathematicians,* Springer, 2003.

[22] G. Tononi, O. Sporns, G. Edelman, A measure for brain complexity: relating functional segregation and integration in the nervous system, *Proc. Natl. Acad. Sci. USA,* **91** (1994), 5033–5037.

[23] G. Tononi, O. Sporns, G. Edelman, A complexity measure for selective matching of signals by the brain, *Proc. Natl. Acad. Sci. USA,* **93** (1996), 3422–3427.

[24] G. Tononi, O. Sporns, G. Edelman, Measures of degeneracy and redundancy in biological networks, *Proc. Natl. Acad. Sci. USA,* **96** (1999), 3257–3262.

Laboratoire de Mathématique d'Orsay - C.N.R.S. (U.M.R. 8628) & Université Paris-Sud, Université Paris-Sud, F-91405 Orsay Cedex, France

E-mail address: jerome.buzzi@math.u-psud.fr

Laboratoire de Probabilités et Modèles Aléatoires (CNRS U.M.R. 7599) and Université Paris 6 – Pierre et Marie Curie, U.F.R. Mathematiques, Case 188, 4 place Jussieu, 75252 Paris cedex 05, France

E-mail address: lorenzo.zambotti@upmc.fr